ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Brain and Language xxx (2003) xxx–xxx

Brain and Language

www.elsevier.com/locate/b&l

Characterizing the time course of an implicature: An evoked potentials study

Ira A. Noveck* and Andres Posada

Institut des Sciences Cognitives, 67 blvd. Pinel, 69675 Bron Cedex, France

Accepted 21 February 2003

Abstract

This work employs Evoked Potential techniques as 19 participants are confronted with sentences that have the potential to produce scalar implicatures, like in *Some elephants have trunks*. Such an Underinformative utterance is of interest to pragmatists because it can be considered to have two different truth values. It can be considered true when taken at face value but false if one were to treat *Some* with the implicature *Not All*. Two accounts of implicature production are compared. The neo-Gricean approach (e.g., Levinson, 2000) assumes that implicatures intrude automatically on the semantics of a term like *Some*. Relevance Theory (Sperber & Wilson, 1985/1996) assumes that implicatures are effortful and not automatic. In this experiment, the participants are presented with 25 Underinformative sentences along with 25 sentences that are Patently True (e.g. *Some houses have bricks*) and 25 that are Patently False (e.g. *Some crows have radios*). As reported in an earlier study (Noveck, 2001), Underinformative sentences prompt strong individual differences. Seven participants here responded true to all (or nearly all) of the Underinformative sentences and the remaining 12 responded false to all (or nearly all) of them. The present study showed that those who responded false to the Underinformative sentences took significantly longer to do so that those who responded true. The ERP data indicate that: (a) the Patently True and Patently False sentences prompt steeper N400's—indicating greater semantic integration—than the Underinformative sentences and that (b) *regardless of one's ultimate response* to the Underinformative sentences, the N400's were remarkably flat, indicating no particular reaction to these sentences. Collectively, the data are taken to show that implicatures are part of a late-arriving, effort-demanding decision process.

Keywords: Linguistic-Pragmatics; N400; Implicature; Scalar terms

1. Introduction

A large number of studies have employed ERP techniques to investigate semantic and syntactic aspects of sentence processing. These studies typically present specific anomalies in a sentence in order to capture a characteristic pattern that follows. Kutas and Hillyard (1980a, 1980b) pointed out how semantic anomalies give rise to a central parietal negative-going component that peaks about 400 ms after the appearance of an inappropriate word, like *socks* in (1); this is known as an N400. The word is not semantically associated with the rest of the sentence nor could one argue that it is anticipated. An ungrammatical structure gives rise to a late centroparietal positivity around 600 ms after this word's onset (this is known as a P600). For example, the word *to* in (2) points to such an anomaly.[1] In contrast, it is more difficult to study pragmatic anomalies because these often thrive on the anomalousness of the sentence itself. Consider (3) below:

(1) John buttered his bread with *socks*.
(2) The broker *persuaded to* sell the stock.
(3) *Some elephants have trunks.*

Syntactically and semantically, the sentence in (3) is correct and, taken quite literally, it is obviously true. We know that elephants in general have trunks, from which it logically follows that (at least) some of them do. What

* Corresponding author. Fax: +33-437-911-210.
E-mail address: noveck@isc.cnrs.fr (I.A. Noveck).

[1] As Osterhout and Holcomb (1995, p. 194) point out, the verb persuade, in an active form, does not allow for a prepositional phrase or an infinitival clause to occur immediately adjacent to the verb.

0093-934X/03/$ - see front matter
doi:10.1016/S0093-934X(03)00053-1

might make the sentence seem odd, is not the syntax or the semantics but the pragmatic fact that it is much less informative than common knowledge would allow. Where the sentence says "some," "all" would be more appropriate. According to standard pragmatic views, when a speaker utters a relatively weak term (e.g. *Some*) on a scale of informativeness, it is an indication that she chose not to use a more informative term from the same scale (e.g. *All*).[2] She thereby conveys that the stronger term *All* is not applicable in the context, (or at least that she does not know whether it is). Thus, uttering *Some* implicates *Not All* (and *Not All* is logically equivalent to *Some are not*). This can lead one to interpret (3) as meaning also *Some elephants do not have trunks*. This kind of inference has been dubbed *scalar implicature* and has since become a paradigmatic case for the study of implicature in the linguistic-pragmatic literature. Understood as carrying a scalar implicature, a sentence like (3) is not true but false.

The application of ERP techniques to implicatures is useful because it could help resolve an ongoing debate in the linguistic-pragmatic literature. Although linguistic-pragmatists agree on the output of the implicature process, they differ with respect to its *automaticity*. One school of thought, which we will refer to as the "neo-Gricean" account (Levinson, 2000), assumes that while the logical interpretation is the more basic one (i.e. the one assigned to *Some* by grammar), the pragmatic interpretation is actually the *default interpretation* in concrete communicative situations. That is, such an interpretation tends to occur (as a consequence of the implicature) every time *Some* is encountered; cases of logical interpretations are those in which the implicature is undone by the context. A second approach, Relevance Theory, does not assume the implicature is automatic but that it is produced when searching for a relevant interpretation of an utterance (Carston, 1999; Sperber & Wilson, 1985/1996). Thus, Relevance Theory considers implicature an effortful, non-necessary inference, whereas, acording to the neo-Griceans it should be the occasional undoing of such an implicature that takes extra effort.

Here, we follow up on a previous investigation that focused on establishing the reality of scalar implicature in everyday reasoning by showing how implicatures became evident with age in standard developmental tasks (Noveck, 2001). That study employed five exemplars of the sort presented in (3) and showed that, whereas a significant majority of linguistically competent children tend to treat the sentence as true, adult participants tend to be equivocal between true and false. Roughly 33% of adult participants considered all five such items true and 40% false The remaining participants tended to consider the items false by indicating that 3 or 4 of the five were false. (From here on we will refer to the true responses as "logical" and the false responses as "pragmatic.") That work presented the two theoretical accounts but was not designed to determine which of the approaches was better supported. The present work aims to adjudicate between the two accounts. It does so by presenting a longer series of sentences like (3), exclusively to adults, along with control items that are Patently True or Patently False, like in (4) and (5), respectively:

(4) Some houses have bricks.
(5) Some crows have radios.

One can consider two kinds of dependent measures that can help reveal the way implicatures are processed here—reaction times and ERP's. With respect to reaction time data, the neo-Gricean approach and Relevance Theory approach make opposing predictions. If the neo-Gricean approach is correct, logical responses to items like (3) ought to take longer than pragmatic responses because it is assumed that the implicature arrives by default and that the logical response is the due to a supplementary step of undoing the implicature. From a Relevance Theory perspective, the initial interpretation, taken at face value, should correspond to the true response and an implicature ought to prompt further responding and a false response. So, logical responses to these Underinformative sentences will be faster than pragmatic responses.

The literature reveals some indirect evidence for claims from a Relevance Theory perspective. Rips (1975) showed that participants take less time to evaluate categorization items like *Some congressmen are politicians* when instructions indicate that *Some* ought to be interpreted as *Some and perhaps all* than when the instructions indicate that *Some* ought to be interpreted as *Some but not all*. That study, however, used more complex materials, made comparisons across two somewhat different experiments, and did not allow for spontaneous interpretations of *Some*, as we do here.

With respect to ERP's, the items exemplified by (4) and (5) can act as benchmarks for determining the way participants treat the items exemplified by (3). On the one hand, if ERP profiles of participants for items like (3) resemble those found in responding to (4) it would be an indication that scalar implicatures are treated like Patently True items. On the other hand, if participants' ERP profiles of items like (3) resemble those found in responding false to (5), it would indeed be an indication that scalar implicatures are automatic or early in sentence processing. The second outcome would be in line with neo-Gricean account according to which an im-

[2] The scale of informativeness can be determined by entailment relations. The stronger term entails the weaker but not vice versa. All is a stronger quantifier than Some because All entails Some while Some does not entail All (to say that All Italians like ice-cream logically implies that Some Italians like ice-cream; however, Some Italians like ice-cream does not necessarily imply that All Italians like ice-cream).

plicature is automatic and intrudes on the semantic interpretation process. Moreover, anticipated individual differences can be investigated to see how those participants who respond true and those who respond false to items like in (3) compare; Perhaps the "logical" responders to (3) elicit profiles like those for (4) and "pragmatic" responders those for (5). In short, we seek evidence indicating that the production of implicatures occurs on line and we do so by determining whether Underinformative sentences provoke N400 responses like the control problems.

The N400 literature can actually gain from this investigation because the prior literature prompts one to make two opposing predictions with respect to sentences like (3). On the one hand, the ERP literature indicates that *contextual constraint* plays a determining role in prompting relatively large N400's; the less anticipated the word, the larger the N400 (see Coulson, 2001). Based on this description of the N400, a sentence like (3) might lead to a relatively large N400 because the sentence's final word (being so obviously related to elephants) is arguably unexpected. A relatively large N400 to (3), with respect to sentences like those in (4), would be an indication that the detection of the pragmatic anomalousness of such a sentence is made on-line. Of the two opposing claims being investigated here, a relatively large N400 would have to be viewed as more favorable to the neo-Gricean approach because it would indicate that some sort of pragmatic process is intruding on the semantic integration of the sentence as the last word appears (and that that prompts the scalar implicature or is perhaps due to it).[3]

On the other hand, existing ERP literature on sentence verification tasks indicates that *mismatching* among items mentioned in the sentence appears to be the cause of larger N400's after the last word rather than the judged truth or falsity of the sentence. For example, it has been shown that items like *A sparrow is a bird* and *A sparrow is not a bird* lead to comparable ERP's while items like *A sparrow is a tool* or *A sparrow is not a tool* prompt greater negativity than the two mentioned earlier (Fischler, Childers, Achariyapaopan, & Perry, 1985). Similarly, Kounios and Holcomb (1992) claimed that quantifiers do not affect N400's even when the truth conditions of items vary as a result. For example, *All apples are fruit* and *No apples are fruits* have comparable effects on the amplitudes of N400's (see Kutas & Van Petten, 1994, for a brief summary). Kounious and Holcomb's study also employed the quantifier *Some* in its category judgement tasks, but it instructed participants to respond true to items like *Some apples are fruits* thus deflecting the potential false response that indicates that an implicature had reared its head. The present study is a more severe test of Kounios and Holcomb's claim because, aside from employing materials that are arguably easier than those in categorisation judgements, it allows for both interpretations of *Some* in Underinformative items and, as a result, allows for variability in truth-judgment across participants to the *same* sentence. If indeed different judgments to the same sentence do not affect ERP waveforms that would be even stronger evidence that truth judgments do not matter to N400 measures. Also, we will focus on the group of participants who responded false to the Underinformative items. This way, one can determine the extent to which the response ("false") is linked to the N400. More importantly to the theoretical interests of the present work, if sentence (3) prompts a relatively small N400 with respect to, say, (5), regardless of one's ultimate response, it would indicate that nothing occurs automatically in such pragmatically anomalous sentences and that scalar implicatures are more likely the product of a post-semantic decision process.

2. Materials and methods

2.1. Pretest

We first replicated the findings from Noveck (2001) with a pencil-and-paper test using a more extended set of materials that included 25 Patently True items, 25 Patently False items, and 25 Underinformative items, i.e., they can be judged to be true or false depending on a logical or pragmatic interpretation. We purposely used 25 items of each type because we wanted to eliminate or reduce the possibility of rote learning; in the meanwhile—even with an eventual artifact removal—we were sure to have enough data in order to make valid comparisons.

Twenty-three Masters students of educational psychology in Grenoble, France were asked to evaluate the materials on one of two dimensions (in French). One group of 11 students was asked to indicate "which of the following would be better said with the word All" and another group of 12 was asked to indicate whether each of the items is true or false. The items were presented in one of two random orders.

For the "better said with … All" group, there was high agreement. Both the Patently True and Patently False items generally were not considered candidates for such a designation. For any given Underinformative item, on average 10.0 of 11participants indicated that it was "better said with… All." Most of the 25 were identified by 10 or all 11 participants; in the worst case, an Underinformative item was identified by 7 participants. For the "true/false" group, those items designed to be Patently True or Patently False led to over-

[3] In fact, the neo-Gricean approach does not make this prediction specifically, but such an outcome would arguably be more favorable to it than to the one based on Relevance Theory.

whelming agreement. As expected, the Underinformative items led 5 participants to consider (24.4 of) the 25 items true and 7 to consider (23.9 of) them false. As these numbers indicate, the results were polarized. No one was equivocal within their task. Overall percentages are in keeping with those found by Noveck (2001) with just 5 items of each. The 75 items, shown to produce the desired response, were employed in the ERP study below.

2.2. Participants

Participants were 19 volunteers who work or study at the Institut des Sciences Cognitives. All were right-handed with normal or corrected-to-normal vision. They were all native speakers of French (age range, 21–32, eight were men).

2.3. Procedure

The study was conducted in French. The participants were given instructions on paper that said that the experiment concerned the comprehension of the word *Some* (*Certains* in French). They were told that they would be given brief items one word at a time and that it was their task to indicate whether the presented sentence was true or false by hitting the appropriate buttons. Three original examples were given on paper (one example representing each condition), with true and false indicated next to each, in order to give the participants a clearer idea of the items that would appear. The examples were *Some men have computers* (Patently True), *Some dogs have ears* (Underinformative), *Some crows have radios* (Patently False). Participants were told that, prior to each sentence, there would be a plus sign (“+”) presented in order to focus their attention on the center of the screen. It was there that the words appeared one at a time, centered in the screen for 200 ms with an interword interval of 40 ms.

2.4. EEG recording

The electroencephalograms (EEG) was recorded with a 65 channel Geodesic Sensor Net through AC-coupled high input impedance amplifiers (200 kΩ, Net Amps). Amplified analogue voltages (0.1–200 Hz bandpass) were sampled at 500 Hz. Electrode impedance was kept below 50 kΩ. ERP analyses consisted in averaging the EEG segments in synchronization with the onset of the last word in each trial over a 1100 ms period including a 100 ms pre-stimulus interval. The signals were low-pass filtered (25 Hz) and a baseline correction was calculated from the 100 ms pre-stimulus interval and the signal was re-referenced using average-reference. Twelve standard 10-10 system scalp sites were determined according to the 64 Geodesic Sensor Net electrode position (Luu & Ferree, 2000). Four central sites were used—midline frontal (Fz), central (Cz), parietal (Pz), and occipital (Oz)—along with lateral pairs of electrodes over frontal (F3 and F4), central (C3 and C4), parietal (P3 and P4), and parietal–occipital (PO3 and PO4). Trials contaminated by eye blinks or eye movements were not included in the analysis (which accounted for 13.6% of the ERP data).

2.5. Data analysis

Reaction times were assessed and ERP's to the last word of each sentence were recorded.[4] Both data sets used within-subject analyses of variance (ANOVA's) with participants as repeated measures. Independent variables were Utterance type (Patently True vs. Patently False vs. Underinformative) and, when necessary, Responder type (logical vs. pragmatic). ERP analyses employed two levels of Electrode Site as factors (Latitudinal: Left, Midline, and Right; Longitudinal: Frontal, Central, Parietal, and Occipital–Parietal) as repeated measures. The dependent measure was the average voltage amplitude in the 300–500 ms latency range. We carried out a global N400 analysis and analyses based on individual differences. Significant interactions were followed up with simple effects analyses.[5]

3. Results

We first analyzed the data from the reaction times. Reaction times that took longer than 3000 ms were removed from these analyses (this accounted for 2.38% of the data). Rates of correct responses were high to the Patently True and Patently False items: 95.5% correctly endorsed the Patently True items and 98.1% correctly rejected the Patently False items. We now focus our attention on the Underinformative items, which are the ones designed to provoke an implicature.

Table 1 summarizes the Reaction Time data. In keeping with the paper-and-pencil test, there were marked individual differences indicating internal consistency. Each participant provided responses to the Underinformative items that were, by a significant majority, either consistently “true” or consistently “false.” Out of the 19 participants, seven (37%) consistently responded true to the Underinformative items (these participants respond true to the Underinformative

[4] Underinformativeness can only be determined upon arriving at the last word; nevertheless, N400's on the word Some were also investigated revealing nothing of interest based on Responder Type nor anything else.

[5] Analyses using time windows of 100 ms (250–350, 350–450, and 450–550) and that investigated pairs of electrodes (F3/F4, etc.) and the midline were also conducted; however, these revealed nothing novel and are thus not reported.

YBRLN 2765 ARTICLE IN PRESS No. of pages: 8; 4C: 6
DISK / 19/3/03 / Bush(CE) / Manju(TE) DTD 4.3.1 / SPS

Table 1
Reaction times to the presented items as a function of response type to the Underinformative statement

Type of response to the Underinformative statement	Patently True statement	Patently False statement	Underinformative
Those who respond logically to the Underinformative statement ($n = 7$)	647	633	655
Those who respond pragmatically to the Underinformative statement ($n = 12$)	1064	856	1203
Total	911	774	1014

Note. Those who respond logically to the Underinformative items (e.g., *Some elephants have trunks*) choose true and those who respond pragmatically choose false (see text for explanation).

statement at a rate that averages 96%). The remaining 12 (63%) consistently responded false to the Underinformative items (at a rate that averages 92%).

On the basis of this individual difference, we carried out a 2×3 ANOVA in which the two kinds of Responses to the Underinformative statement (what we refer to as *Logical* and *Pragmatic* responders) were a between-participant variable and the three Utterance types (Patently True, Patently False, and Underinformative) were a within-participant variable. Results revealed that the two kinds of responders prompted significantly different reaction times overall, $F(1,17) = 19.008$, $p < .0005$. Utterance type also yielded significant differences, $F(2,34) = 13.209$, $p < .0001$. Individual comparisons by t test (and using .0166 as the level of significance) revealed that the Patently False items yielded significantly faster response times than both the Patently True items, $t(18) = 3.245$, $p < .005$ and the Underinformative items $t(18) = 5.38$, $p < .0001$; furthermore, the difference between the Patently True items and the Underinformative ones was marginally significant $t(18) = 2.299$, $p = .0337$. The ANOVA also revealed a significant interaction, $F(2,34) = 7.153$, $p < .005$. The interaction is due, at least partly, to the fact that the gap between Pragmatic and Logical responders is wider among the Underinformative items than it is for the Patently False items.

4. ERP analysis

4.1. N400

The grand average ERP's in the three conditions is depicted in Fig. 1. All three conditions yielded similar waveforms and there was a steeper negativity in the right hemisphere. The ANOVA of the 300–500 ms latency range took Utterance type (Patently True vs. Patently False vs. Underinformative) vs. Laterality (right vs. midline vs. left) vs. Anterior–Posterior (AP) location (four levels) as within-participant factors and yielded a main effect of Utterance type $F(2,36) = 4.858$, $p < .05$. There were no significant interactions between Utterance type and AP and were no significant interactions between Utterance type and Laterality. There was a main effect of Laterality $F(2,36) = 3.567$, $p < .05$, indicating greater activity in the right hemisphere of the scalp when compared to the left, and a main effect of AP, $F(3,54) = 9.265$, $p < .0001$, indicating far greater negativity in the posterior portions of the scalp than in the anterior portions. There was an interaction between Laterality and AP, indicating greater negativity in the posterior portion of the Right hemisphere $F(6,108) = 4.078$, $p = .001$.

Pairwise comparisons showed that the Patently False items differed significantly from the Underinformative items, $F(1,18) = 5.906$, $p < .05$. More interestingly, this analysis revealed that Patently True items yielded significantly larger N400s than the Underinformative ones as well, $F(1,18) = 9.561$, $p < .01$ and that the Patently True and Patently False items did not significantly differ from each other $F(1,19) = .597$, $p = .45$ The increased negativity among the Patently True and Patently False items indicate that these items prompt significantly more semantic integration than the Underinformative items; the waveforms from the Underinformative items indicate that they required little semantic integration overall.

4.2. N400's and individual differences

We now shine a spotlight on the 12 pragmatic responders. It is important to determine whether those who respond false to the Underinformative utterances reveal any evidence of contextual integration or not. That is, do the N400's of the Pragmatic responders show any sign of on-line pragmatic intrusion on the semantic integration of Underinformative utterances?

An ANOVA for the 12 Pragmatic responders group was undertaken by employing the same three factors as before: Utterance type (Patently True vs. Patently False vs. Underinformative) vs. Laterality (right vs. middle vs. left) vs. Anterior–Posterior (AP) location (four levels). It yielded a main effect of Utterance type $F(2,22) = 7.272$, $p < .005$. This analysis also led to a significant interaction between Utterance type and AP (which we will ad-

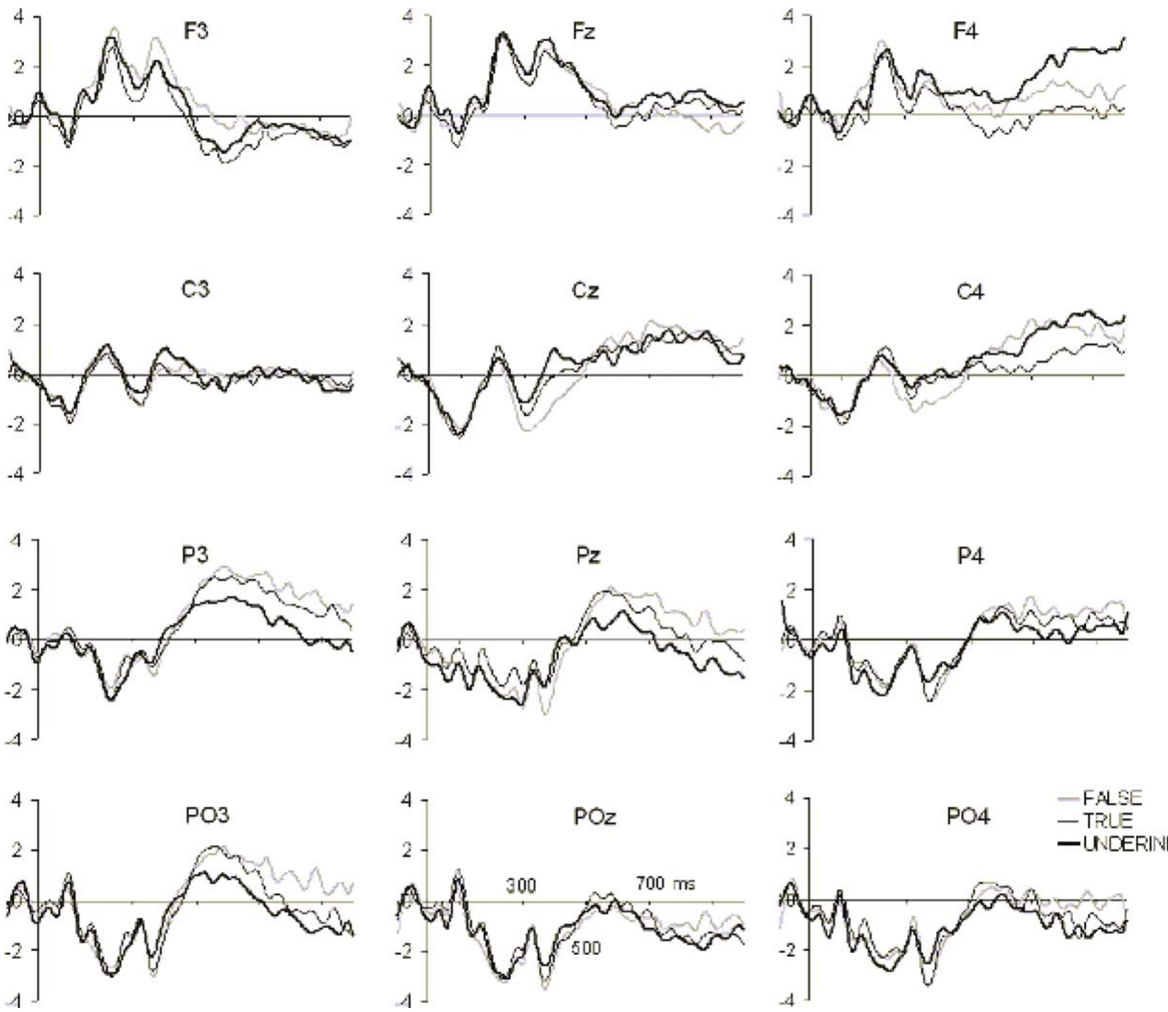

Fig. 1. The grand average ERP's in the three conditions—Patently False, Patently True, and Underinformative.

dress directly below). Pairwise comparisons among these Pragmatic responders pointed to the same differences described in the initial global (N400) ANOVA, and largely accounts for the significant results in that ANOVA: Patently true items among the pragmatic responders yielded significantly larger N400s than the Underinformative items, $F(1,11) = 6.102$, $p < .05$. Patently False items, too, yielded significantly larger N400s than the Underinformative items, $F(1,11) = 9.538$, $p < .05$. As before, the Patently True and Patently False items are statistically comparable $F(1,11) = 3.777$, $p = .08$, though here one sees that the Patently False items nearly lead to N400's that are more extreme than those prompted by the Patently True ones. This is what is captured by the AP $\times$ Utterance interaction reported above. The Patently False items among the Pragmatic responders tended to prompt steeper negativity than the two other Utterance types in the posterior section of the scalp.

To verify that Responder Type had no effect on the Underinformative items, we carried out an ANOVA of the 300–500 ms latency range with respect to the Underinformative items. The 2 (Responder Type: Logical vs. Pragmatic) $\times$ 3 Laterality (right vs. midline vs. left) $\times$ 4 Anterior–Posterior (AP) location ANOVA with the last two being within-participant factors, yielded no main effect for Responder Type, $F(1,17) = 0.009$, $p > .92$, and there were no significant interactions involving Responder Type.

In sum, it is noteworthy that the Pragmatic responders' reactions to the Underinformative items (i.e., their N400's) remain unremarkable when compared to the Patently True and the Patently False items, indicating little evidence that they require much semantic integration. This is despite the fact that they respond false. Moreover, there is no difference at all between the Logical and the Pragmatic responders to the Underinformative items.

5. Discussion

This work has been concerned with characterizing the responses to Underinformative sentences like *Some ele-*

phants have trunks. The reaction-time data are particularly striking for three reasons. First, the difference in reaction time between those who respond true to the Underinformative statements and those who respond false is large (655 and 1203 ms, respectively). False responses to the Underinformative statement take nearly twice as long as the true responses. Second, there is spillover. Those who respond true to the Underinformative items are also significantly faster in responding correctly to the two other conditions. Thus, the individual difference reflects two sorts of strategies. Those who respond true to the Underinformative items are responding literally and quickly overall and those who respond false are responding non-literally and slowly overall. Third, it is important to keep in mind that the Patently False items yielded the fastest response times overall. Thus, the relative slowness of the *false* responders to the Underinformative items occurs despite evidence of preparedness for the Patently False items. This finding makes it difficult to argue that Underinformative items, by representing one-third of the stimuli, allowed for a rote response among the pragmatic responders. In sum, indications from the behavioral data are that those who give a false response to the Underinformative items undertake deeper processing that is, in turn, evident in the responses to the other items in the task. The deeper processing linked to the false responses in the Underinformative condition does not conform to expectations based on the neo-Gricean account, but it does with Relevance Theory because it assumes that implicature production arrives as a result of an effortful process.

The ERP data were especially instructive. The N400 data indicate that the Underinformative items prompt little semantic integration at all and less so than *both* the Patently True and Patently False items. The Underinformative items generally lead to flat N400's. This is also clearly seen for those who were *pragmatic* in their responses, indicating that deeper processing was not linked to any particular activity concerning the N400.

The ERP findings are actually highly reasonable in light of previous work. Kutas and colleagues indicate that N400's either capture associativity between words or else the likelihood that a particular word is anticipated (Kutas & Hillyard, 1980a, 1980b,1984; van Petten & Kutas, 1991). They argue that N400's are steeper when a target word (which would be the final word here) (a) is not associated with the prior context or; (b) is just unanticipated. The Underinformative items in this study allow one to distinguish between these two accounts. The final words for the Underinformative items here are arguably as unanticipated as the final words in the two other conditions, and yet one finds no real signs of semantic integration for them. At the same time, the Underinformative sentences contain final words that are by definition highly associated with the subject of the sentence. Thus, the present work points to *disassociativity* among words as the more likely of the two factors to predict N400's. This would explain why the N400 linked to the final word of a sentence like *Some elephants have trunks* is linked with little semantic integration while sentences like *Some houses have bricks* and *Some birds have televisions* reveal significantly more. A trunk is a more essential feature of elephants than bricks are to houses or birds are to televisions.

Kounios and Holcomb's findings—that truth-judgements do not matter as much as associativity—can be said to have been confirmed and extended. All participants show no particular reaction to the Underinformative utterances and even the pragmatic (false) responders to items like (3) ultimate produce flat N400s. This would indicate that semantic features of the final word (with respect to the subject of its sentence) is the variable factor that reflects the steepness of the N400.

Most importantly to our theoretical aims, the fact that the ERP profiles for the pragmatic group of participants in the Underinformative condition remains unremarkable, even as their responses and response times indicate much deliberation, is further evidence that participants' immediate reaction to the Underinformative statement is a benign one. This indicates that the hypothesized implicature, which prompts participants to respond false, arrives at a later stage and eventually requires more effort than the responses that prompt a true response. Indications are then that the false responses to the Underinformative sentences (which are arguably prompted by the implicature) are effort demanding and late arriving; that is, the false response appears linked to decision-related mechanisms that arguably arrive after those indexed by the N400 (see Heinze, Muente, & Kutas, 1998, for a similar argument with respect to a categorization paradigm). Thus, the findings here conform with existing claims in the ERP literature

The findings are also consistent with claims based on Relevance Theory. In the Relevance framework, an implicature is defined as an inference that the speaker *intends* and *expects* the hearer to draw in order to arrive at an interpretation of the utterance that is relevant enough. In particular, a scalar implicature is derived when a relatively weak statement fails to meet the hearer's expectation of relevance. For instance, in the dialogue:

(6) Isaac: Do all Italians like ice cream?
Noemi: Some do.

Noemi's answer is not relevant enough unless it is taken to implicate that *Some Italians do not like ice cream*. For Relevance theory, scalar implicatures are derived only when they are contextually needed to achieve the expected level of relevance and the relevant interpretation of an utterance is determined by the listener's attempt to gain as many effects as possible for the least effort. Given that a scalar implicature is an inference that goes further than the semantic origins of words like *Some* and that it is carried out to achieve relevance, it is not entirely surprising that we have found evidence showing that not all participants provoke a scalar implicature and that it is a late occurring, effortful step for those who choose to do so.

Acknowledgments

This work was supported by a grant from the CNRS (ATIPE, 1999) to the first author. The authors wish to express their gratitude to Shlomo Bentin, Lewis Bott, Seanna Coulson, Michel Hoen, Guy Politzer, Dan Sperber, and an anonymous reviewer whose comments improved the paper.

Appendix A

Ten examples of the Patently True, Patently False, and Underinformative items (translated from French).

Patently True	Patently False	Underinformative
Some people have brothers	Some couches have windows	Some turtles have shells
Some animals have stripes	Some cars have parents	Some giraffes have necks
Some houses have bricks	Some kangaroos have airplanes	Some sentences have words
Some flags have stars	Some fruit have computers	Some cherries have pits
Some buildings have elevators	Some dogs have wings	Some staircases have steps
Some rugs have stains	Some ducks have cassettes	Some televisions have screens
Some teeth have cavities	Some shoes have hats	Some books have pages
Some shirts have buttons	Some crayons have pants	Some beaches have water
Some houses have garages	Some birds have televisions	Some airplanes have wings
Some gardens have trees	Some toads have churches	Some dogs have ears

References

Carston, R. (1999). Informativeness, relevance and scalar implicature. In R. Carston, & S. Uchida (Eds.), *Relevance theory: Applications and implications*. Amsterdam: John Benjamins.

Coulson, S. (2001). *Semantic leaps: Frame shifting and conceptual blending in meaning construction*. Cambridge: Cambridge University Press.

Fischler, I., Childers, D. G., Achariyapaopan, T., & Perry, N. W. (1985). Brain potentials during sentence verification: Automatic aspects of comprehension. *Psychophysiology, 20*(4), 400–409.

Heinze, H. J., Muente, T. F., & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology, 47*, 121–135.

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*(4), 459–479.

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

Kutas, M., & Hillyard, S. A. (1980b). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Psycholinguistics* (pp. 83–133). San Diego: Academic Press.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Luu, P., & Ferree, T. (2000). Determination of the Geodesic sensor net's average electrode positions and their 10-10 international equivalents. Electrical Geodesics, Inc., Technical Note, pp. 1–15.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*(2), 165–188.

Osterhout, L., & Holcomb, P. J. (1995). Event related potentials and language comprehension. In M. D. Rugg, & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event related brain potentials and cognition: Vol. 25* (pp. 171–215). New York: Oxford University Press.

Rips, L. J. (1975). Quantification and semantic memory. *Cognitive Psychology*.

Sperber, D., & Wilson, D. (1985/1996). *Relevance: Communication and cognition*. Oxford: Basil Blackwell.

van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open and closed class words. *Memory and Cognition, 18*, 380–393.