



Contents lists available at ScienceDirect

## Journal of Memory and Language

journal homepage: [www.elsevier.com/locate/jml](http://www.elsevier.com/locate/jml)

## On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities

Mante S. Nieuwland<sup>a,b,c,\*</sup>, Tali Ditman<sup>b,c</sup>, Gina R. Kuperberg<sup>b,c,d</sup>

<sup>a</sup> Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd Floor, 20009 Donostia, Spain

<sup>b</sup> Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA

<sup>c</sup> MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, Bldg 149, 13th Street, Charlestown, MA 02129, USA

<sup>d</sup> Department of Psychiatry, Massachusetts General Hospital, Bldg 149, 13th Street, Charlestown, MA 02129, USA

### ARTICLE INFO

#### Article history:

Received 19 August 2009  
revision received 22 June 2010  
Available online 15 July 2010

#### Keywords:

Language comprehension  
Pragmatics  
Scalar quantifiers  
Electrophysiology  
N400 ERP  
Individual differences  
Pragmatic abilities  
Autism Quotient Communication subscale  
Informativeness

### ABSTRACT

In two event-related potential (ERP) experiments, we determined to what extent Grice's maxim of informativeness as well as pragmatic ability contributes to the incremental build-up of sentence meaning, by examining the impact of underinformative versus informative scalar statements (e.g. "Some people have lungs/pets, and...") on the N400 event-related potential (ERP), an electrophysiological index of semantic processing. In Experiment 1, only pragmatically skilled participants (as indexed by the Autism Quotient Communication subscale) showed a larger N400 to underinformative statements. In Experiment 2, this effect disappeared when the critical words were unfocused so that the local underinformativeness went unnoticed (e.g., "Some people have lungs that..."). Our results suggest that, while pragmatic scalar meaning can incrementally contribute to sentence comprehension, this contribution is dependent on contextual factors, whether these are derived from individual pragmatic abilities or the overall experimental context.

© 2010 Elsevier Inc. All rights reserved.

### Introduction

According one of the key principles of pragmatics, addressees by default presume that speakers communicate efficiently by uttering messages that are informative (Grice, 1975; Sperber & Wilson, 1986). This so-called *conversational maxim of quantity* is based on the idea that communication has evolved as a cooperative effort, and it often implicitly shapes our communicative interactions (e.g., Engelhardt, Bailey, & Ferreira, 2006; see also Clark, 1996). Of course, that does not mean that everything that we say or write is genuinely informative. We easily adjust our expectations to who we are talking to (e.g., children, people who know more or

less than we do), reflecting the fact that what is informative or relevant to one individual might be trivial or irrelevant to another. Moreover, there is abundant literature to suggest that individuals can vary greatly in their abilities to produce and comprehend pragmatic language, which could mean that some people are simply more focused on the logic of utterances than others (e.g., Baron-Cohen, 2008).

Although Grice's account of pragmatic principles was not intended to serve as a psychological model of cognitive processing (see Bach, 2006; Bezuidenhout & Cutting, 2002), it may be that the addressee's default presumptions have important ramifications for how language is processed on-line (e.g., Wilson & Sperber, 2004). One way in which Grice's maxim of quantity may play out in on-line sentence processing is by influencing the addressee's expectations of what kind of words will come next (e.g., Federmeier, 2007; Van Berkum, 2009). For example, following the sentence fragment "Some people have...", the addressee might expect the upcoming word to denote something that not all

\* Corresponding author at: Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 2nd Floor, 20009 Donostia, Spain. Fax: +34 943 309 052.

E-mail addresses: [m.nieuwland@bcbi.eu](mailto:m.nieuwland@bcbi.eu), [mante@nmr.mgh.harvard.edu](mailto:mante@nmr.mgh.harvard.edu) (M.S. Nieuwland).

people have (e.g., ‘pets’, ‘tattoos’), instead of something that all people possess (e.g., ‘lungs’, ‘bodies’). As a result, one can hypothesize that trivially true, underinformative statements (e.g., “Some people have lungs”) incur semantic processing costs because they deviate from the addressee’s expectations. In the two experiments reported below, we determined to what extent Grice’s maxim of informativeness contributes to the incremental build-up of sentence meaning. Specifically, we explored differences in individual’s reliance on this maxim for interpretation, and also investigated the role of general contextual factors on the processing of underinformative utterances. We addressed these issues by examining the impact of underinformative versus informative scalar sentences (e.g. “Some people have lungs/pets...”) on the N400 event-related potential (ERP), an electrophysiological index of semantic processing (Kutas & Hillyard, 1980, 1984).

Ever since Aristotle’s science of logic, quantifiers and logical operators have been important windows into human reasoning, and have maintained a crucial role in logic and linguistics because of their association with truth-value (e.g., see Gamut, 1991). The scalar quantifier ‘some’ has received much attention because it allows for two disparate readings: a pragmatic interpretation and a logical interpretation. The pragmatic interpretation approximates to ‘some but not all’ or ‘only some’. This interpretation constitutes a *conversational inference*, by which language comprehenders attribute an implicit meaning beyond the logical or literal meaning. This inference is termed a *scalar inference* or *scalar implicature* because it is thought that comprehenders base this pragmatic interpretation on the assumption that the communicator had a reason for not using a more informative or *stronger* term on the same quantity scale (some < many < all; see Horn, 1972). In other words, comprehenders assume that the communicator would have said ‘all’ if he/she thought ‘all’ was true, and assume that the communicator says ‘some’ because he/she thinks that stronger expressions like ‘many’ and ‘all’ are false.

The logical interpretation approximates to ‘at least some’ or ‘some and possibly all’. This interpretation makes sense when communicators use the expression ‘some’ when they lack all the relevant information (for example, “Some guests are coming to my party, but not everybody has RSVPed yet”, in which case it is possible that many or all invitees will come to the party), or when they are not referring to a specific subset (e.g., “Some people were crossing the street”).

Importantly, the pragmatic and logical interpretation may yield different truth-values. For a simple, informative statement like “Some people have pets”, each interpretation yields an outcome that is true with respect to world knowledge; it is true that ‘some but not all people have pets’, consistent with the pragmatic interpretation, and it is also true that there exist people with pets, consistent with the logical interpretation. However, for an underinformative statement like “Some people have lungs”, whereas the logical interpretation yields a true outcome (because people with lungs do exist), the pragmatic interpretation yields a false outcome (because all people have lungs, not just some). The fact that ‘some’ may yield disparate truth-values can be used to examine how language comprehenders apply their pragmatic knowledge during

sentence comprehension and establish sentence truth-value (for reviews see Noveck and Reboul (2008), Noveck and Sperber (2007) and Sedivy (2007)).

Theoretical accounts of how people deal with scalar quantifiers predominantly differ in whether they assume that scalar inferences are generated by default or whether scalar inferences are context-dependent (see also Geurts, 2009; Horn, 2006; Recanati, 2003). In what has been dubbed the Levinsonian account, scalar inferences are generated automatically upon encountering ‘some’. The idea behind this is that, because the pragmatic meaning of scalars is so dominant in our language use, it has become ‘lexicalized’ (see Levinson, 2000; for related accounts see Chierchia, 2004; Gazdar, 1979) such that the intended message can be efficiently communicated. The pragmatic meaning, however, can be cancelled when the subsequent context requires so. For example, upon encountering the sentence “John wanted some of the cookies”, addressees automatically generate the pragmatic interpretation and interpret the sentences as meaning John wanted some, but not all, of the cookies. However, at a later point, upon encountering the sentence “In fact, he wanted all of them”, they revise their initial interpretation to be consistent with the logical interpretation. According to this account, it is this undoing of the scalar inference that is costly.

In contrast, proponents of Relevance Theory have posited that the generation of scalar inferences is chiefly a function of whether the inference is required to meet the addressee’s standard of relevance (e.g., Carston, 1998; Sperber & Wilson, 1986). The logical interpretation of ‘some’ (i.e., “some and possibly all”) could very well lead to a satisfying interpretation of the utterance, but the discourse context may require the addressee to derive a scalar inference to arrive at the pragmatic interpretation. Since this pragmatic interpretation involves ‘narrowing’ (negation of the stronger expressions ‘many’ and ‘all’), it constitutes a fully fledged inferential process which requires processing time and effort beyond the ‘easier’ logical interpretation.

Neither the Levinsonian framework nor Relevance Theory constitutes a psychological model of scalar inferences with explicit implications for processing. Yet, experimental psychologists have tried to infer testable predictions about the time course of scalar inferences. It has been argued that if scalar inferences are generated automatically, as advocated in the Levinsonian account, they are also generated relatively rapidly and their cancellation would incur additional processing costs (e.g., Bott & Noveck, 2004). In contrast, if scalar inferences are truly context-dependent, then they would incur processing costs in situations where they are not licensed by the context. According to Breheny, Katsos, and Williams (2006), Relevance Theory predicts that in a neutral context (i.e., without a discourse context that biases towards either a logical or a pragmatic interpretation), no scalar inference will initially be computed, and only when the logical interpretation is deemed insufficient will addressees invest additional cognitive effort to generate a scalar inference.

To examine the time course for the generation of scalar inferences, behavioral research on scalar inferences has often used the sentence-verification paradigm (e.g., Bott &

Noveck, 2004; De Neys & Schaeken, 2007; Feeney, Craffton, Duckworth, & Handley, 2004; Noveck, 2001; Noveck & Posada, 2003; Pijnacker, Hagoort, Buitelaar, Teunisse, & Geurts, 2009; for reviews, see Bezuidenhout and Morris (2004), Huang and Snedeker (2009), Noveck and Reboul (2008), Noveck and Sperber (2004, 2007) and Sedivy (2007)). In sentence-verification tasks participants are asked to judge the truth of a statement, and in speeded sentence-verification tasks participants are asked to do this as fast as possible. Because the logical and pragmatic interpretation of informative sentences yield identical truth-values, the dependent measure of whether a scalar inference has been made is whether participants respond 'false' to an underinformative scalar statement (e.g., "Some people have lungs"). An often reported finding is that participants who respond 'false' to underinformative sentences are slower than those who respond 'true' (e.g., Bott & Noveck, 2004; Noveck & Posada, 2003; Rips, 1975). This is the case regardless of whether participants are explicitly instructed to respond 'false' or whether they spontaneously decide to do so (e.g., Bott & Noveck, 2004). These results have been interpreted as suggesting that scalar inferences are associated with additional processing costs and result from a delayed decision process (e.g., Bott & Noveck, 2004; Noveck & Posada, 2003; Noveck & Reboul, 2008).

Although using a sentence-verification task makes intuitive sense when dealing with truth-value, its interpretation is subject to a number of important caveats, as has already been noted by several researchers (Feeney et al., 2004; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2009). For example, evaluating the logical meaning of an underinformative sentence may be inherently easier than evaluating its pragmatic meaning because one needs only one or two examples to verify the logical meaning (one or two people that have lungs) whereas one may need to do a more extended analysis to falsify the pragmatic meaning (e.g., search of, and failing to find counterexamples in memory; see also Grodner et al., 2010; Huang & Snedeker, 2009). Thus, it may not necessarily be the case that *generating* the pragmatic meaning requires additional processing effort and time, but rather *refuting* it. Another important concern is that speeded sentence verification is a relatively unnatural task that may encourage participants to ignore their pragmatic knowledge (Feeney et al., 2004), and it is hardly representative of how people process language in everyday life. Importantly, people who do generate scalar inferences are also slower in other conditions (e.g., Noveck & Posada, 2003), suggestive of a more general difference in task-related strategic processing. Finally, reaction times in verification tasks are generally quite slow, over 600 ms when statements are presented word-by-word (e.g., Bott & Noveck, 2004; Noveck & Posada, 2003) or even in the order of seconds when sentences are presented as a whole (e.g., De Neys & Schaeken, 2007; Pijnacker et al., 2009). In this regard, the results from verification tasks should be taken to reflect early stages of language processing as well as the output of downstream decision processes that follow them (e.g. Kounios & Holcomb, 1992).

Recently, researchers have overcome these problems by using a more indirect, high temporal resolution measure of

scalar processing – the visual-world paradigm. Using this paradigm, Huang and Snedeker (2009) recorded eye-movements while participants received auditory instructions such as "Click on the girl that has some of the socks" or "Click on the girl that has all of the soccer balls" in the presence of a display in which one girl had two socks from the four socks that were present in the display, and another girl had all three soccer balls that were present in the display. The temporary referential ambiguity in the instruction at the point of 'some' could, in principle, be resolved immediately if participants made a scalar inference that would restrict 'some' to a proper subset. Participants, however, were substantially delayed, to 'some', but not when the instruction contained the word 'all'. Based on this observation, Huang and Snedeker argued that 'pragmatic' scalar inferences are delayed relative to the 'semantic' logical interpretation (see also Bott & Noveck, 2004; Breheny et al., 2006; De Neys & Schaeken, 2007; Noveck & Posada, 2003).

However, Grodner and colleagues (Grodner et al., 2010) note that 'some' is not unambiguously associated with a scalar inference (e.g., "Click on the girl with some socks" does *not* imply other socks are in the discourse), and that it was the partitive construction 'of the' that allowed for identification of the target in the Huang and Snedeker study. In contrast, for *all*, the quantifier itself was sufficient to identify the target. In a related study by Grodner et al. (2010) that circumvented these and some additional issues, scalar inference associated with pragmatic-*some* was not delayed relative to expressions that did not require a scalar inference. Thus, in contrast to the Huang and Snedeker (2009) results, the Grodner et al. results suggest that the pragmatic meaning of scalar expressions is rapidly available.

In the present study on scalar processing, we employed another indirect, high temporal resolution measure of language comprehension, namely Event-Related Potentials (ERPs). An important advantage of ERPs is that they provide both quantitative and qualitative information about language processing well in advance of (and without the principled need for) an explicit behavioral response (e.g., Van Berkum, 2004). In particular, we focus on the N400 ERP component (Kutas & Hillyard, 1980, 1984; see Kutas, Van Petten, and Kluender (2006), for review), a negative deflection in the ERP that emerges somewhere between 150 and 300 ms after the onset of a word and that peaks at about 400 ms, with a maximum over the back of the head (i.e., electrodes at parietal locations). The N400 is, in principle, elicited by every content word, and its amplitude decreases in size and in a gradual manner when the word fits the context better (e.g., Kutas et al., 2006; Van Berkum, Brown, Zwitterlood, Kooijman, & Hagoort, 2005). A differential effect of two conditions on the N400 amplitude is referred to as an N400 effect. The functional significance of the N400 is still under debate (e.g., Kutas et al., 2006; Lau, Phillips, & Poeppel, 2008; Van Berkum, 2009), but there is a general consensus that its amplitude reflects the fit between the lexical-semantic meaning of an incoming word and the interaction between linguistic context (at the level of single words, sentences and discourse) with information stored in memory (e.g., semantic memory,

real-world knowledge and pragmatic knowledge of what a speaker is likely to say), henceforth referred to as ‘semantic fit’.<sup>1</sup> The results from recent ERP studies have shown that the interaction between context and real-world knowledge can lead people to generate expectations about the semantic properties of specific upcoming words (e.g., DeLong, Urbach, & Kutas, 2005; Federmeier, 2007; Van Berkum, 2009; Van Berkum et al., 2005), although it may be that, under other circumstances, the three-way mapping process is initiated only once the word is encountered. Importantly, in a recent study on negation processing, we showed that the N400 ERP is also sensitive to the informativeness of an utterance (Nieuwland & Kuperberg, 2008). In this study, participants read sentences that were true but underinformative due to pragmatically unlicensed negation (e.g., “Bulletproof vests aren’t very *dangerous*...”, in which case negation is used to deny something that makes no sense to begin with, namely that bulletproof vests *are* dangerous). Critical words (‘dangerous’) in these sentences elicited an increased N400 responses in the same way that false sentences did. In contrast, true sentences that contained pragmatically licensed negation (e.g., “With proper equipment, scuba-diving isn’t very dangerous...”) elicited N400 responses that were indistinguishable from those elicited by true affirmative sentences (e.g., “With proper equipment, scuba-diving is very safe...”). These results suggest that pragmatic knowledge of what is an informative thing to say influences an early stage of semantic processing, and may even contribute to building up broad pragmatic expectancies about what upcoming words are likely to be encountered.

There has been one previous study investigating whether the N400 is modulated by scalar inferences. Noveck and Posada (2003) recorded readers’ electrophysiological responses to sentence-final words in underinformative sentences (e.g., “Some elephants have *trunks*”), patently false sentences (e.g., “Some crows have *radios*”) and patently true sentences (e.g., “Some houses have *bricks*”). Similar to previous behavioral studies, participants were asked to make a sentence verification response following each sentence. The results indicated that patently true and patently false sentences elicited a larger N400 ERP than underinformative sentences, and that the N400 responses to underinformative sentences were not modulated by whether participants responded true or false to these sentences. Consistent with previous behavioral findings, the reaction time data indicated that those participants who made scalar inferences (i.e. responded ‘false’ to underinformative sentences) were much slower to respond than those who followed a literal interpretation (i.e., responded ‘true’ to underinformative sentences). Critically, however, participants who made scalar inferences were much slower in all conditions, suggesting that these participants were using a more cautious strategy overall (see Feeney et al., 2004, for a related discussion). Noveck and Posada interpreted the smaller N400 for underinformative sentences, in combination with the slow time course of

scalar implicatures, as being inconsistent with a Levinsonian account. They also suggested that scalar implicatures may likely be the product of a post-semantic decision process, that, once the critical word has been encountered, computes the truth-value of the complete proposition, whereas the initial stage of semantic processing after the critical word is determined only by simple lexical-semantic relationships (e.g., see also Fischler, Bloom, Childers, Roucos, & Perry, 1983; Kounios & Holcomb, 1992). Later accounts by Noveck and colleagues suggest that, under certain conditions, the pragmatic scalar meaning may be generated without having to traverse through a logical interpretation first (Noveck & Sperber, 2007). However, the general idea that pragmatic processing costs are incurred after lexico-semantic processing is complete has persisted in some models of language processing (e.g. Bornkessel-Schlesewsky & Schlewsky, 2008; Cutler & Clifton, 1999; Fodor, 1983; Forster, 1979; Regel, Gunter, & Friederici, 2010).

Several problems with interpreting the initial ERP study by Noveck and Posada. First, the materials in the different conditions were not matched or counterbalanced, and the words were presented in at a very fast pace (a presentation duration of 200 ms per word and an inter-word interval of 40 ms, which is about half of what is customarily used in ERP research using serial visual presentation).<sup>2</sup> Second, they employed a sentence-verification task that may have evoked decision-related positive ERPs that overlap in time and scalp distribution with the N400, and that may obscure modulations of the N400 (e.g., Kuperberg, 2007). In light of these concerns, it is important to note that patently false sentences did not evoke larger N400 responses than patently true sentences, whereas violations of real-world knowledge have consistently been associated with larger N400 responses in other studies (e.g., Fischler, Childers, Acharyapapan, & Perry, 1985; Fischler et al., 1983; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Hald, Steenbeek-Planting, & Hagoort, 2007; Nieuwland & Kuperberg, 2008). This is problematic because these violations were included to establish a benchmark comparison for the main results.

In the current study, we addressed some of these concerns and used ERPs to examine how rapidly different individuals use their pragmatic knowledge of what is an informative versus uninformative thing to say during the processing of scalar sentences. We compared ERP responses elicited by critical words in underinformative scalar statements (e.g., “Some people have lungs, ...”) to those elicited by critical words in informative scalar statements (e.g., “Some people have pets, ...”, see Table 1 for more examples). If the pragmatic meaning of weak scalar quantifiers can be used incrementally during sentence comprehension (i.e., scalar inferences are made online), this may

<sup>1</sup> This view can be distinguished from one in which the N400 reflects the combinatorial process of integrating a critical word with the preceding context or of assessing the plausibility of the resulting proposition (see Kuperberg, 2007; Lau et al., 2008; Van Berkum, 2009, for discussion).

<sup>2</sup> The short presentation duration that was used by Noveck and Posada (and by Bott and Noveck (2004)), although constant, may mimic the speed of the natural reading rate more closely. However, using these durations in the RSVP procedure, which does not allow backtracking or slowing down, can cause readers to experience difficulties with normal sentence comprehension (see Camblin, Ledoux, Boudewyn, Gordon, & Swaab, 2007), and note that word-by-word self-paced reading times are generally over at least 350 ms even for very short words (see Koornneef & Van Berkum, 2006; Ditman, Holcomb, & Kuperberg, 2007a).



**Table 1**

Examples sentences from Experiment 1. Critical words are underlined for expository purpose only.

Underinformative/informative
Some people have <u>lungs/pets</u> , which require good care
Some rock bands have <u>musicians/groupies</u> , sometimes with drug problems
Some gangs have <u>members/initiations</u> , and often strong hierarchy too
Relatively poor/good semantic fit
Literature classes sometimes read <u>papers/poems</u> as a class
Wine and spirits contain <u>sugar/alcohol</u> in different amounts
Fillers
Many people catch the flu, especially in the winter
Many vegetarians eat bean curd, which is rich in protein

guide expectations about upcoming words so that readers and listeners will expect new input to be informative (e.g., Altmann & Steedman, 1988; Crain & Steedman, 1985; Tanenhaus & Trueswell, 1995; see also MacDonald, Pearlmuter, & Seidenberg, 1994). Given that the N400 is sensitive to how well a word fits the context based on both semantic and pragmatic constraints (Coulson, 2004; Kutas et al., 2006; Nieuwland & Kuperberg, 2008; Van Berkum, 2009; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008), this incremental account predicts that critical words in an underinformative statement would yield a larger N400 than in an informative statement.

In contrast, if the pragmatic meaning of weak scalar quantifiers is not readily available when readers encounter the critical word, then the N400 ERP would not be sensitive to whether the statement is informative or underinformative. Rather, sentence processing and modulation of the N400 may be driven purely by lexico-semantic relationships (e.g., Otten & Van Berkum, 2007; Van Petten, 1993; Van Petten, Weckerly, Mclsaac, & Kutas, 1997; for review, see Kutas et al. (2006)). Because critical words in the underinformative condition (e.g., 'lungs') had a stronger lexical-semantic relationship to the main noun phrase in the preceding phrase (e.g., 'people') than in the informative condition (supported by their higher values on a Latent Semantic Analysis, LSA Landauer & Dumais, 1997), see "Methods"), this would predict a smaller N400 to informative than non-informative sentences (as shown by Noveck & Posada, 2003). This prediction also follows from Grice's original account (for discussion see Geurts, 2009), and is generally consistent with models of language comprehension that assume that pragmatic factors come into play after an initial stage of 'context-free', linguistic-semantic processing (e.g. Fodor, 1983; Forster, 1979).

Previous studies have reported that individuals can vary significantly in whether and how they apply their pragmatic knowledge (e.g., Joliffe & Baron-Cohen, 1999; Musolino & Lidz, 2006; Noveck, 2001; Schindele, Lüdtke, & Kaup, 2008; Stanovich, & West, 2000; Tager-Flusberg, 1981). Moreover, there have been several reports of individual differences in scalar inference generation (e.g., Bott & Noveck, 2004; Fee-ney et al., 2004; Noveck & Posada, 2003), suggesting that different people may preferentially and consistently adopt either a literal or a pragmatic interpretation when asked to evaluate underinformative sentences. Our hypothesis,

which we will describe in more detail below, is that individuals with good real-world pragmatic skills are, at least initially, relatively more sensitive to the pragmatic 'violation' of underinformativeness and therefore more likely to show a pragmatic N400 effect, whereas processing in people with poorer real-world pragmatic skills is more likely to be driven by pure lexico-semantic association.

As a caveat, inferences regarding the full extent of incremental scalar processing based on our paradigm are limited. As opposed to studies that have used the visual-world paradigm, our study was not designed to examine whether scalar inferences are generated immediately upon encountering the scalar quantifiers. A modulation of the N400 by informativeness in our study could be taken either as evidence that the processing consequences of the scalar quantifier either are rapidly computed upon encountering the critical word, or were perhaps computed before encountering the critical word. As argued by Van Berkum (2009), there are good reasons to assume that a pragmatic modulation of the N400 does not directly reflect a fully compositional enrichment process, but more likely indicates that the semantic and pragmatic consequences of the preceding discourse have been computed to serve as an interpretive background to retrieve word meaning (see also Kuperberg, Paczynski, & Ditman, 2010b). Although our study was not specifically designed to examine ERP responses to scalar quantifiers, we will report exploratory analyses that address these issues.

## Experiment 1

In the first of our two experiments, we examined electrophysiological responses to critical words in underinformative statements versus informative scalar statements, and used this measure to investigate individual differences in pragmatic processing. If scalar pragmatic inferences are generated incrementally during on-line sentence processing, critical words that render a statement trivial or underinformative should lead to additional semantic processing costs, and should elicit a larger N400 than critical words in informative statements – a pragmatic N400 effect (see also Nieuwland & Kuperberg, 2008). If, on the other hand, pragmatic scalar information is not used incrementally during online processing, the N400 should not be larger to critical words in underinformative statements. In fact, given the closer lexico-semantic associations in underinformative than in informative sentences (people-lungs versus people-pets), the N400 may even be relatively attenuated in underinformative sentences.

We also hypothesized that there may be individual variation in these patterns of N400 modulation, that may be predicted by variation in participants' abilities to produce and comprehend pragmatic aspects of language in the real world (e.g., Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Happé, 1993; Schindele et al., 2008; Tager-Flusberg, 1981, 1985). We therefore obtained an independent measure of pragmatic language abilities of our participants in everyday life through the Communication subscale of the Autism-Spectrum Quotient questionnaire (the AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) that quantifies an individual's pragmatic skills on a

continuum from autism to typicality. Of the five AQ subscales, the Communication subscale taps into pragmatic abilities most directly. Some examples of items from this subscale are “Other people frequently tell me that what I’ve said is impolite, even though I think it is polite”, “I find it hard to ‘read between the lines’ when someone is talking to me”, and “I am often the last to understand the point of a joke”.

We predicted that individuals with good pragmatic abilities (as indexed by a low score on the AQ-Communication subscale), would be relatively more sensitive to the pragmatic ‘violation’ of underinformativeness and more likely to show a pragmatic N400 effect, as compared to less pragmatically skilled individuals (see Pijnacker et al., 2009; Schindele et al., 2008, for related hypotheses in participants with high-functioning autism or Asperger’s syndrome). This sensitivity may play out in several different ways. For example, individuals with good pragmatic abilities might generate pragmatic inferences more consistently, generate more robust inferences, they might be better at evaluating incoming words for informativeness, or perhaps even have a different task set than people with poor pragmatic skills. In the current study we cannot distinguish between these or other possibilities. Nevertheless, modulation of a pragmatic N400 effect by pragmatic abilities could provide evidence that such everyday communication problems may be, in part, driven by an impaired incremental use of pragmatic knowledge during language processing.

In order to examine the specificity of these potential individual differences, we also included sentences that did not contain scalars, but that contained a word that had a relatively good semantic fit versus relatively poor semantic fit to the preceding sentence context based on real-world knowledge; see Table 1 for examples). We predicted that words that were incongruous with real-world knowledge<sup>3</sup> would produce a robust N400 effect compared to words that were congruous with real-world knowledge (e.g., Kutas & Hillyard, 1984) in all individuals, regardless of their AQ-Communication scores. This allowed us to dissociate individual differences in incrementally recruiting pragmatic knowledge from the more general recruitment of real-world knowledge during online processing.

## Methods

### Participants

Thirty-one right-handed Tufts students (17 males; mean age = 20.2 years) gave written informed consent. All were native English speakers, without neurological or psychiatric disorders.

### Materials

We constructed 70 sentence pairs such that the underinformative and informative versions of each sentence pair

were identical except for the critical word. Each sentence consisted of two clauses, and the first clause (the quantifier clause) always started with the quantifier ‘some’ and always ended with a comma after the critical word. We selected critical words so that replacing ‘some’ by the quantifier ‘all’ would yield a true statement in the underinformative condition (e.g., “All people have lungs”), but a false statement in the informative condition (e.g., “All people have pets”). The second clause always contained at least three words and provided additional information about the critical word, the main NP in the scalar clause (e.g., ‘people’) or the scalar clause as a whole, and was created so that the complete sentence constituted a logically true statement in each condition. Critical words in the two conditions were approximately matched for average length in number of letters (underinformative, informative,  $M = 6.7/7.0$ ,  $SD = 1.8/2.0$ ) and log frequency (Kucera & Francis, 1967; underinformative, informative,  $M = 1.73/1.91$ ,  $SD = 2.29/2.03$ ). Semantic similarity values were calculated for the critical words within the underinformative and informative sentences using Latent Semantic Analysis (Landauer & Dumais, 1997; Landauer, Foltz, & Dumais, 1998; available on the Internet at <http://lsa.colorado.edu>). As expected, underinformative words yielded a higher LSA value than informative words (underinformative, informative;  $M = .33/.17$ ,  $SD = .23/.18$ ;  $t(138) = 4.58$ ,  $p < .001$ ). As noted in the Introduction, higher LSA values are generally associated with smaller N400 amplitudes compared to lower LSA values, because the LSA values reflect in part the amount of lexico-semantic priming a word receives from the preceding context.

For the semantic fit manipulation, we constructed another 70 sentence pairs that were identical except for the critical word. Critical words were selected that were relatively congruous or incongruous to the sentence with regard to world knowledge (see Table 1 for examples). Critical words in the two conditions were matched for average length in letters (congruous, incongruous,  $M = 6.4/6.3$ ,  $SD = 2.1/1.7$ ) and log frequency (Francis & Kucera, 1982; congruous, incongruous,  $M = 1.46/1.50$ ,  $SD = 1.74/1.88$ ). Semantic similarity values were calculated for the congruous, incongruous words using Latent Semantic Analysis. Good semantic fit words yielded a higher LSA value than poor semantic fit words (congruous, incongruous,  $M = .22/.14$ ,  $SD = .11/.08$ ;  $t(138) = 5.31$ ,  $p < .001$ ). At least two words followed the critical words before the sentence ended.

We also created 35 filler sentences that each had a similar sentence structure as the scalar sentences but that always started with the quantifier ‘many’, and involved a simple and true statement (e.g., “Many vegetarians eat bean curd, which is rich in protein.”).

We created two counterbalanced lists so that each sentence appeared in only one condition per list, but in all conditions equally often across lists. Within each list, items were pseudorandomly mixed with the 70 sentences containing a semantic fit manipulation (35 containing a relatively good fitting critical word, 35 containing a relatively poor fitting critical word) and the 35 filler sentences to limit the succession of identical sentence types, while matching trial-types on average list position.

<sup>3</sup> We use the term ‘incongruous with real-world knowledge’, but these sentences did not describe events that are impossible in the real-world, and this term only refers to the relative poor fit with real-world knowledge compared to the congruous sentences.

### The Autism-Spectrum Quotient

The AQ (Baron-Cohen et al., 2001) is a self-administered questionnaire that is designed to measure the extent to which adults with normal intelligence possess traits associated with Autism Spectrum Disorder (ASD). Although this scale is not a diagnostic measure, its discriminative validity as a screening tool has been clinically tested (Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen, 2005). The test consists of 50 items, made up of 10 questions assessing five subscales: Social Skill (e.g., “I would rather go to a library than a party”), Communication (e.g., “I frequently find that I don’t know how to keep a conversation going”), Imagination (e.g., “When I’m reading a story, I find it difficult to work out the characters’ intentions”), Attention To Detail (e.g., “I usually notice car number plates or similar strings of information”), and Attention-Switching (e.g., “I frequently get so absorbed in one thing that I lose sight of other things”). Half the questions are worded to elicit an ‘agree’ response and the other half a ‘disagree’ response, addressing demonstrated areas of cognitive characteristics in ASD (American Psychiatric Association, 1994; Baron-Cohen et al., 2001). Higher scores on the AQ indicate stronger presence of traits associated with ASD. A score of 32+ appears to be a useful cutoff for distinguishing individuals who have clinically significant levels of autistic traits (Baron-Cohen et al., 2001; the maximum score of the participants in our study was 30). Such a high score on the AQ however does not mean that an individual has autism, because a diagnosis is only merited, based on diagnostic measures such as the American Psychiatric Association (1994), ADI-R (Lord, Rutter, & Le Couteur, 1994) or ADOS-G (Lord, Risi, Lambrecht, et al., 2000), if the individual is suffering a clinical level of distress as a result of their autistic traits. In the current study, the AQ was administered in a quiet room subsequently to the ERP experiment, and took each participant about 10 min.

### Procedure

Participants silently read sentences, presented word-by-word and centered on a computer monitor, while minimizing eye-movements and blinks. There was no task other than reading for comprehension. To parallel natural reading times (Legge, Ahn, Klitz, & Luebker, 1997), all words were presented using a variable presentation procedure (Otten & Van Berkum, 2008; see also Nieuwland & Kuperberg, 2008). Word duration in ms was computed as ((number of letters × 27) + 187), with a 10 letter maximum. Also, to mimic natural reading times at clause boundaries (e.g., Hirotani, Frazier, & Rayner, 2006; Legge et al., 1997; Rayner, Kambe, & Duffy, 2000), critical words (which were followed by a comma) were presented for an additional 227 ms, and sentence-final words for an additional 500 ms. All inter-word-intervals were 121 ms. Following sentence-final words, a blank screen was presented for 500 ms, followed by a fixation mark at which subjects could blink and self-pace onto the next sentence by a right-hand button press. Participants were given six short breaks. Total time-on-task was approximately 40 min. After the ERP experiment, each subject was allowed a short break to wash up and was then

administered a brief exit-interview, followed by the Autism-Spectrum Quotient questionnaire.

In the exit-interview, participants received a booklet that contained 6 pages and were instructed to answer the question from the booklet page-by-page without looking at the subsequent pages. On page 1, subjects were asked to report whether they noticed anything about the sentences they read and what research question(s) they thought the experiment was about. On page 2, an example of an informative scalar sentence was given, and participants reported whether they thought that sentences starting with ‘Some’ stood out, what they thought the purpose of these sentences was, and what research question these sentences involved. On page 3, subjects reported whether they thought that some of the sentences in the experiment sounded odd and provided a brief explanation why they thought this. On pages 4 and 5, subjects were presented with 10 different scalar statements, including informative and underinformative scalar sentences truncated after the CW as well as longer sentences that contained locally informative or underinformative phrases. Subjects were asked to rate whether each sentence was true (1 = false, 5 = true) and how normal they would find it if somebody said this (1 = odd, 5 = normal). On page 6, subjects were informed that a sentences like “Some people have lungs” could be rated as false (because the sentence implies that most people do not have lungs) or true (because there are at least some people in the world that do have lungs). The subjects were asked to report whether they thought *during the experiment* about whether these sentences were true or false, whether they *during the experiment* ‘treated’ these sentences as true or false, and how consistently they did this (1 = very inconsistently, 5 = very consistently).

### EEG recording

The electroencephalogram (EEG) was recorded from 29 tin electrodes held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH, USA). Electrode locations included Fz, Cz, Pz, Oz, Fp1/2, F3/4, F7/8, FC1/2, FC5/6, C3/4, T3/4, T5/6, CP1/2, CP5/6, P3/4, P7/8, O1/2, and 2 additional EOG electrodes; all were referenced to the left mastoid). The EEG recordings were amplified (band-pass filtered at 0.01–40 Hz) and digitized at 200 Hz. Impedance was kept below 5 k $\Omega$  for EEG electrodes. Prior to off-line averaging, single-trial waveforms were automatically screened for amplifier blocking and muscle/blink/eye-movement artifacts over 850 ms epochs (starting 100 ms before CW onset). Two participants were excluded due to excessive artifacts (mean trial loss >50%). For the remaining 29 participants, average ERPs (normalized by subtraction to a 100 ms pre-stimulus baseline) were computed over artifact-free trials for CWs in all conditions (mean trial loss across conditions 11%, range 0–42%, without substantial differences in mean trial loss across conditions).

### Statistical analysis

For all analyses reported below, the Greenhouse/Geisser correction was applied to F tests with more than one

degree of freedom in the numerator. Note that due to the large number of trials needed for averaging in ERPs (which reduces the probability that the results hinge on just a few odd items), statistics are only reported for by subjects analyses, and analyses by items are not included.

## Results

### Main effect of informativeness

Critical words elicited very similar N400 responses in the underinformative and the informative statements (see Fig. 1, left panel). Because modulation of the N400 ERP is generally maximal at posterior electrodes (e.g., Kutas et al., 2006), we divided all electrodes into anterior electrodes (F3/4, F7/8, F9/10, FC1/2, FC5/6, FP1/2, FPz, Fz) and posterior electrodes (Pz, Oz, CP1/2, CP5/6, P3/4, P7/8, O1/2) for subsequent analyses. Using mean amplitude in the 350–450 ms time window, a 2 (informativeness: informative, underinformative)  $\times$  2 (AP distribution: anterior, posterior) repeated measures analysis of variance (ANOVA) revealed that there was no statistically significant difference between the ERP responses to informative and underinformative statements, and no interaction effect between informativeness and AP distribution.

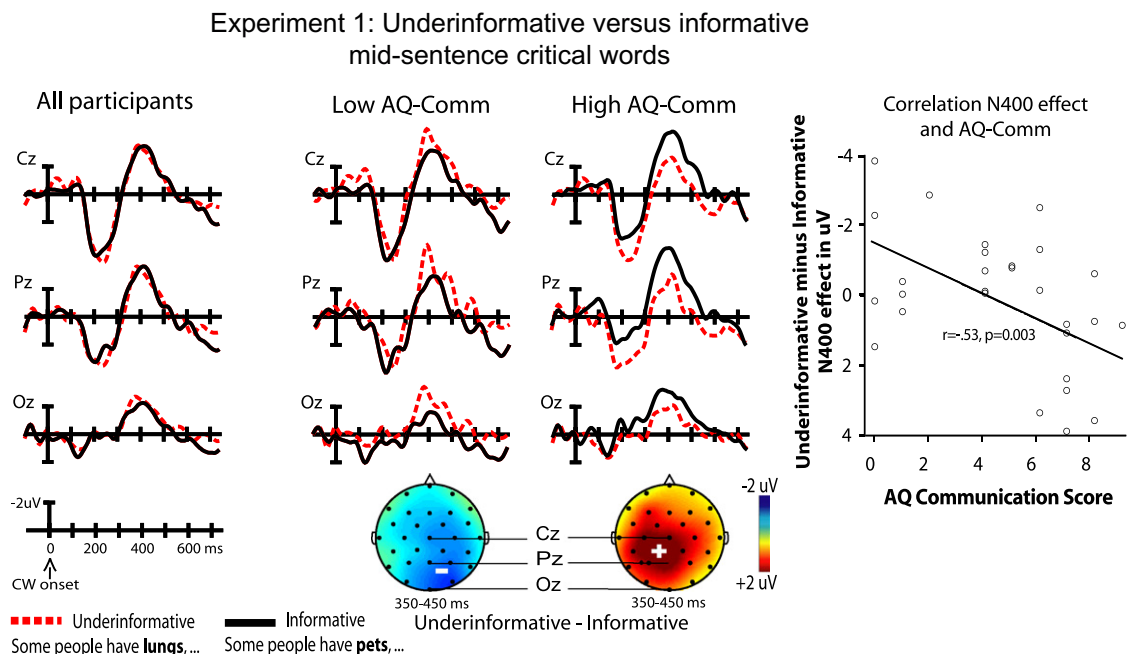
### AQ-Comm score and ERP responses to informativeness

AQ scores ranged from 9 to 30 ( $M = 21$ ,  $SD = 7.04$ ). To explore the role of pragmatic abilities, we first grouped the participants into low AQ-Comm ( $N = 15$ ) and high

AQ-Comm ( $N = 14$ ) groups based on the median split of scores on the Communication subscale. AQ-Comm score for the low AQ-Comm group ranged from 0 to 5 ( $M = 2.33$ ,  $SD = .51$ ; seven males and eight females, mean age 20.9 years, mean total AQ score 15.8), and from 6 to 9 for the high AQ-Comm group ( $M = 7.2$ ,  $SD = .28$ ; eight males and six females, mean age 19.3 years, mean total AQ score 26.5). The two AQ-Comm groups showed statistically significant differences in AQ-Comm score ( $t(27) = 8.34$ ,  $p < .001$ ) and in total AQ score ( $t(27) = 6.24$ ,  $p < .001$ ), as well as in age ( $t(27) = 2.49$ ,  $p < .05$ ; when entered into the subsequent analyses as a covariate, the factor age, however, did not change the patterns of results).

Grand average ERPs for the two groups are displayed in Fig. 1 (middle panel). Using mean amplitude in the 350–450 ms time window, the overall ANOVA revealed a significant 2 (informativeness: informative, underinformative)  $\times$  2 (AQ-Comm Group: low AQ-Comm, high AQ-Comm) interaction effect when using all electrodes ( $F(1, 27) = 9.45$ ,  $p = .005$ ). There was no significant 3-way interaction with AP distribution ( $F(1, 27) = 2.19$ ,  $p = .15$ ), but the Informativeness by AQ-Comm group interaction effect was statistically significant when using only posterior electrodes ( $F(1, 27) = 11.54$ ,  $p = .002$ ), but only marginally significant when using anterior electrodes ( $F(1, 27) = 3.3$ ,  $p = .07$ ). This predominantly posterior distribution of N400 modulation is consistent with the N400 literature (e.g., Kutas et al., 2006).

Simple main-effect analysis for the groups separately, using posterior electrodes only, showed that underinformative statements elicited larger N400 responses than



**Fig. 1.** Left panel: Grand-average event-related potential (ERP) waveforms elicited by critical words in underinformative (dotted lines) and informative (solid lines) statements from Experiment 1, shown at electrode locations Cz, Pz, and Oz. In this and all following figures, negativity is plotted upwards. Middle panel: Grand average ERPs elicited by critical words in underinformative and informative statements per AQ-Communication group in Experiment 1, and corresponding scalp distributions of the mean difference effect (underinformative minus informative sentences) in the 350–450 ms analysis window. Right panel: Correlation between N400 effect and AQ-Communication score.



informative statements in the low AQ-Comm group ( $F(1, 14) = 5.57, p = .033, CI -.82 \pm .75$ ), whereas informative statements elicited larger N400 responses than underinformative statements in the high AQ-Comm group ( $F(1, 13) = 6.12, p = .028, CI -1.38 \pm 1.2$ ). There was no statistically significant effect of informativeness in the two AQ-Comm groups separately when taking into account anterior electrodes only ( $F_s < 1, n.s.$ ).

As can be seen from Fig. 1, there appeared to be differential effects of informativeness for the two groups before the 350–450 ms time window. We therefore performed additional  $2$  (informativeness: informative, underinformative)  $\times 2$  (AQ-Comm Group: low AQ-Comm, high AQ-Comm) ANOVAs for the 50–150, 150–250 and 250–350 ms time windows. These revealed some significant effects within early time windows (50–150 ms in the low AQ-Comm group, 150–250 ms in the high AQ-Comm group; see Appendix A for full report, which can be found at <http://www.nmr.mgh.harvard.edu/kuperberglab/materials.htm>). We were concerned that these early effects of informativeness reflected an artefactual side effect of dividing subjects on the basis of their AQ-Comm score. It is well-known that with limited numbers of EEG trials going into the average of a single subject, single-subject ERPs constitute unknown mixtures of critical ERP effects and residual EEG background noise which could, in principle, explain the early onset ERP differences. We therefore repeated analyses using a longer, 500 ms pre-CW baseline thus reducing noise in the baseline time window (and consequently, in the post-baseline ERP signal). ERP difference effects that truly are the result of the experimental manipulation should survive this longer baseline analysis. The corresponding figures for these analyses can be found at the website as referenced above. After rebaselining, the early effects in the 50–150 and 150–250 ms windows disappeared but left the main pattern of results in the 250–350 and 350–450 ms windows unchanged (see Appendix A). Additional analyses for the post-450 ms time windows using the original baseline as well as the new baseline can also be found on our website.

#### *Correlation analysis for AQ-Comm scores and ERP responses to informativeness*

We also performed a correlation analysis that took into account the full range in individual AQ-Comm scores, and revealed a negative correlation between AQ-Comm score and the mean ERP difference score calculated as underinformative minus informative in the 350–450 ms time window at posterior electrodes (Pearson's  $r = -.53, p = .003$ ; see Fig. 1, right panel). This correlation effect was also present for total AQ score ( $r = -.55, p = .002$ ), the Social Skill subscale score ( $r = -.45, p = .014$ ) and Attention-Switching subscale score ( $r = -.55, p = .002$ ), but was not significant for scores on the subscales Imagination ( $r = -.21, p = .29$ ) and Attention To Detail ( $r = .17, p = .39$ ). We should note that the Attention-Switching subscale and the Communication subscale were also the strongest interrelated subscales, so the effects of these subscales are hard to tease apart.

#### *ERP responses to informativeness and the role of LSA*

As mentioned in the Introduction, the content words in underinformative statements co-occur in language relatively more frequently than those in the informative statements, as reflected by their differences in LSA values. However, not each underinformative statement from each sentence pair had a larger LSA value than its informative counterpart. This allowed us to separate our items into one set that had a relatively small LSA difference between informative and underinformative sentences (LSA(underinformative–informative),  $M = -0.02, SD = 0.12$ ), and one set that had a relatively large LSA difference across conditions ( $M = 0.34, SD = 0.18$ ). By computing ERPs separately for these two sets for each group, we investigated the effect of informativeness while controlling for lexical-semantic factors.

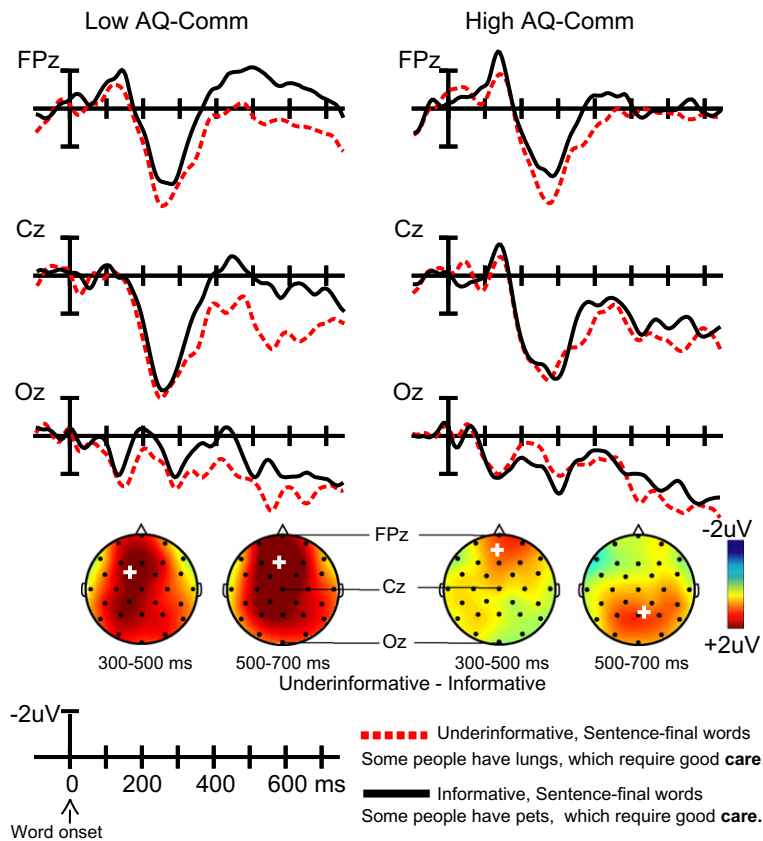
The corresponding figures for these analyses can be found at (<http://www.nmr.mgh.harvard.edu/kuperberglab/materials.htm>). These plots reveal clear differences between the low and high AQ-Comm groups in N400 modulation by LSA and informativeness. Analyses focusing on N400 peak amplitude modulations across posterior electrodes in the 350–450 ms time window showed that the informativeness by LSA difference interaction effect was significant in the high AQ-Comm group ( $F(1, 13) = 5.38, p = .037$ ), but not in the low AQ-Comm group ( $F(1, 14) = .02, p = .90$ ). Follow-ups showed that, in the low AQ-Comm group, critical words in underinformative statements elicited a larger N400 than those in informative statements, both when there was a relatively small and a relatively large LSA difference between conditions (small difference,  $F(1, 14) = 2.37, p = .043, CI -.84 \pm .76$ ; large difference,  $F(1, 14) = 2.19, p = .052, CI -.80 \pm .77$ ). In the high AQ-Comm group, however, underinformative statements elicited a lower N400 than informative statements only when there was a relatively large LSA difference ( $F(1, 13) = 4.01, p = 0.001, CI -2.25 \pm 1.21$ ), but not when there was a relatively small LSA difference ( $F(1, 13) = .21, p = 0.834, CI -.17 \pm 1.76$ ).

In sum, whereas we found a typical modulation of LSA in the high AQ-Comm group, the pragmatic N400 effect in the low AQ-Comm group was insensitive to LSA.

#### *Group differences in ERP responses to sentence-final words*

We also examined the ERP responses to sentence-final words in underinformative and informative statements between the two AQ-Comm groups (see Fig. 2). Statistical analyses were carried out using mean amplitude in the 300–500 ms. The sentence-final words involved different word categories, and there may have been differences in naturalness of the second clauses following informative versus underinformative statement. Our main interest in this comparison was therefore not the main effects of informativeness (positive ERPs to sentence-final words of underinformative than informative sentences across both groups,  $F(1, 27) = 20.28, p < .001, CI .96 \pm .46$ ), but rather the differences between the two AQ-Comm groups to the same set of stimuli. As

### Experiment 1: Underinformative versus informative sentence-final words



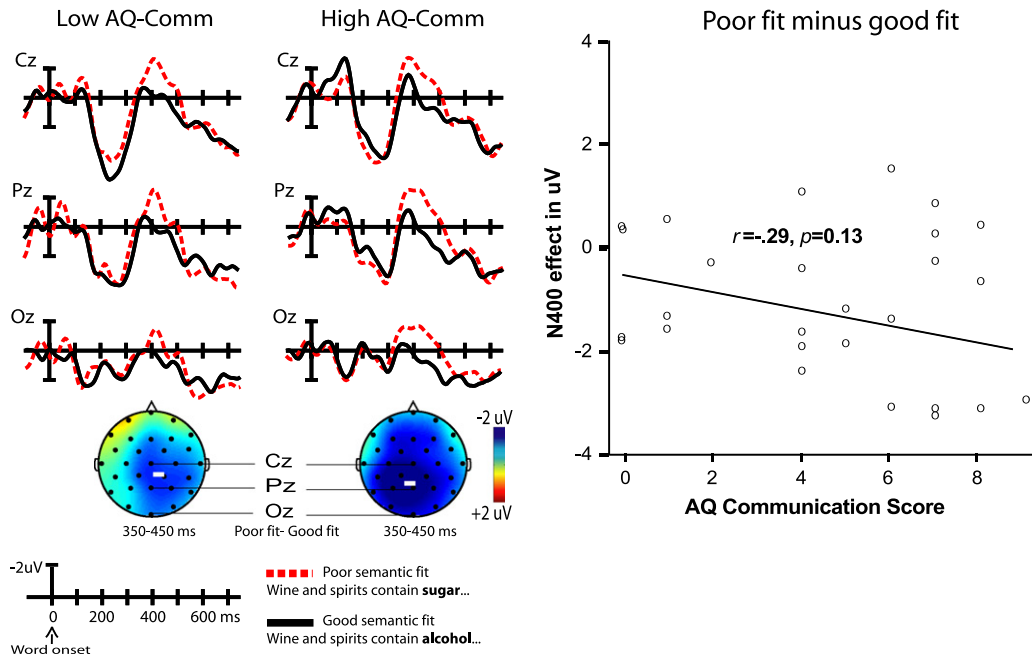
**Fig. 2.** Grand average ERPs elicited by sentence-final words in underinformative (dotted lines) and informative (solid lines) statements per AQ-Communication group in Experiment 1, shown at electrode locations FPz, Cz, and Oz, and corresponding scalp distributions of the mean difference effect (underinformative minus informative sentences) in the 300–500 ms and the 500–700 ms analysis window.

shown in Fig. 2, there was a clear differential ERP effect on the sentence-final words in underinformative and informative statements in the low AQ-Comm group, but less so in the high AQ-Comm group. This differential ERP effect appeared to have a slightly frontal distribution (i.e., inconsistent with an N400 effect scalp distribution), and may reflect additional sentence wrap-up processing. Across all electrodes, the overall ANOVA revealed a marginally significant informativeness by AQ-Comm group interaction effect ( $F(1, 27) = 3.88, p = .059$ ) and follow-ups showed that the modulation by informativeness was significant in the low AQ-Comm group ( $F(1, 27) = 17.56, p = .001, CI 1.36 \pm .70$ ), but only marginally significant in the high AQ-Comm group ( $F(1, 27) = 3.64, p = .079, CI .54 \pm .60$ ). A 2 (informativeness: informative, underinformative)  $\times$  2 (AP distribution: anterior, posterior) ANOVA revealed no interaction effect of informativeness with anterior-posterior distribution ( $F < 1$ ), and there was no significant interaction between informativeness, AQ-Comm group and distribution ( $F < 1$ ). Because the effect was prolonged, we repeated the above analyses in the 500–700 ms window and this yielded the same pattern of results.

#### Group differences in ERP responses to real-world congruous versus incongruous sentences

To determine the specificity of the group differences in ERP responses to underinformativeness, we also examined whether the groups differed in their N400 modulation to words that were congruous versus incongruous with real-world knowledge. We compared the modulation of the N400 by words with a relatively poor versus good fit based on real-world knowledge across the two groups. As can be seen from Fig. 3, the modulation of the N400 was quite similar across the two groups. Using mean amplitude at posterior electrodes in the 350–450 ms time window, the overall 2 (Real-world congruity: congruous, incongruous)  $\times$  2 (AQ-Comm Group: low AQ-Comm, high AQ-Comm) ANOVA revealed that the incongruous words evoked a larger amplitude N400 than congruous words ( $F(1, 27) = 19.28, p < .001, CI -1.35 \pm .64$ ). However, no Real-world congruity by AQ-Comm Group interaction was observed ( $F(1, 27) = 1.77, p = .19$ ). There was also no significant Real-world congruity by AQ-Comm Group interaction in the adjoining 250–350 and 450–550 time windows (all  $F_s < 2, ns.$ ). Consistent with the absence of this interaction, there was also no significant

## Experiment 1: Poor semantic fit versus good semantic fit



**Fig. 3.** Left panel: Grand average ERPs elicited by words that had a relatively poor (dotted lines) and relatively good (solid lines) semantic fit per AQ-Communication group in Experiment 1, and corresponding scalp distributions. Right panel: Correlation between N400 effect and AQ-Communication score.

correlation between the N400 difference effect in the 350–450 ms time window and AQ-Comm score (Pearson's  $r = -.29$ ,  $p = .13$ ).

#### Exploratory analyses of ERP responses to the scalar quantifiers

Although our experiment was not specifically designed to examine ERP responses to the scalar quantifiers, we performed an exploratory analysis to investigate whether there were differences between the two AQ-Comm groups in ERP responses to the sentence-initial scalar quantifiers 'Some' (the sentence-initial word of the experimental sentences) and 'Many' (the sentence-initial word in 35 filler sentences). The reasoning behind this analysis was that if the quantifiers themselves evoke differential pragmatic processing, then the differences in pragmatic abilities between the groups may already become apparent at the quantifier. We note that the quantifier 'many' can elicit a "not all" implicature as can 'some', so this comparison is not optimal for examining differences in pragmatic processing. However, because these quantifiers can be arranged on a scale of informativeness where 'many' is stronger than 'some', the 'some' implicature would include "not many" as well as "not all". In this sense, and particularly in an experimental context in which both are repeatedly presented, one could argue that these scalar quantifiers are associated with implicatures that are of different strength.

The figures corresponding to this analysis can be accessed at <http://www.nmr.mgh.harvard.edu/kuperberg-lab/materials.htm>. In the high AQ-Comm group, 'Many', relative to 'Some' appeared to evoke a slightly more nega-

tive right-lateralized waveform at about 300–350 ms and a more positive frontally-distributed waveform at about 650–700 ms. There appeared to be no such effect in the low AQ group. We performed a series of repeated measures ANOVAs to test for the 2 (quantifier: some, many) by 2 (AQ-Comm group: low AQ-Comm, high AQ-Comm) interaction, in adjoining 50 ms time windows between 100 and 800 ms after quantifier onset, using all electrodes or only anterior or posterior electrodes. The only (marginally) significant interaction effect was found in the 650–700 ms window using anterior electrodes ( $F(1, 27) = 3.77$ ,  $p = .063$ ). Follow-up analyses confirmed that 'many' elicited more positive ERPs than 'some' in the high AQ-Comm group ( $F(1, 13) = 14.70$ ,  $p = .002$ ,  $CI -2.04 \pm 1.15$ ), but there was no difference between the two quantifiers in the low AQ-Comm group ( $F(1, 14) = .07$ ,  $p = .80$ ,  $CI -.21 \pm 1.68$ ). In addition, this frontal positivity effect showed a marginally significant correlation with AQ-Comm score (Pearson's  $r = .34$ ,  $p = .073$ ). There was also a marginally significant correlation between the frontal positivity effect and the differential ERP effect at the critical words, suggesting that participants who showed a larger frontal positive effect were less likely to show a pragmatic N400 effect later in the sentence ( $r = -.35$ ,  $p = .06$ ). The frontal positivity, however, did not predict the N400 modulation by real-world congruity (Pearson's  $r = -.15$ ,  $p = .46$ ).

#### Exit-interview

We examined whether the AQ-Comm groups differed in their exit-interview ratings for truth-value and naturalness. A 2 (AQ-Comm group: low AQ-Comm, high

AQ-Comm) by 2 (informativeness: informative, underinformative) ANOVA revealed no group differences in the truth-value ratings and the naturalness ratings (all  $F_s < 2$ ). In addition, underinformative and informative statements received similar truth-value ratings ( $t < 1$ ) but different naturalness ratings ( $t(1, 28) = 15.98, p < .001$ ).

## Discussion

Across all participants, underinformative statements elicited N400 responses that were similar to those elicited by informative statements. However, there was marked heterogeneity across individuals in N400 modulation, with some individuals showing a larger N400 to critical words in underinformative than in informative statements, and others showing the opposite pattern of modulation (i.e., a larger N400 to critical words in informative than underinformative statements). Most importantly, these individual differences could be explained by taking into account individual variability in real-world pragmatic language ability. Individuals with few pragmatic language difficulties (as indexed by a low score on the AQ-Communication subscale) were more sensitive to the pragmatic ‘violation’ of underinformative statements. This opposite pattern of activity was clear both in a median split analysis that dichotomized the two groups and in a correlation analysis that took into account the full range in individual AQ-Comm scores. Importantly, this N400 modulation by AQ-Comm score did not extend to the N400 responses to words with a relatively poor fit with respect to world knowledge, suggesting that AQ-Comm score was fairly specific in explaining the pattern of N400 modulation to the pragmatic violations. In addition, the two groups were differentially sensitive to lexical–semantic co-occurrence: whereas the low AQ-Comm group showed a pragmatic N400 effect independently of whether the underinformative and informative sentences were matched for LSA, the high AQ-Comm group’s ERP responses were modulated by LSA. Finally, we also explored ERP responses to the scalar quantifier ‘some’ versus ‘many’. Although these quantifiers could be argued to evoke related (although not identical) pragmatic processes, rendering this comparison suboptimal for examining potential differences in pragmatic processing, we did find some preliminary evidence that pragmatic abilities influenced processing at the scalars themselves.

If one considers only the pragmatically skilled participants, our results show that pragmatically underinformative statements are associated with early semantic processing costs (see also Nieuwland & Kuperberg, 2008). This result suggests that the pragmatic meaning of a scalar quantifier can, in principle, be rapidly and incrementally incorporated during sentence comprehension, a finding that is consistent with models of language processing that incorporate an incremental contribution of pragmatic factors (Altmann & Steedman, 1988; Crain & Steedman, 1985; Tanenhaus & Trueswell, 1995) and with the results of studies from the visual-world paradigm (Grodner et al., 2010).

In contrast to the more pragmatically skilled participants, however, the less pragmatically skilled participants

showed no pragmatic N400 effect. Their processing was rather driven primarily by the relatively closer lexical–semantic relationships between individual words in these statements which overrode pragmatic factors. One possible interpretation of these results is that these individuals, who report difficulties with pragmatic abilities in everyday life, were simply incapable of generating scalar inferences. One could argue that this conclusion is in line with the notion from Relevance Theory that scalar inferences are not obligatory (see also Bott & Noveck, 2004; Noveck & Posada, 2003) but depend on constraints from the context and possibly from neuropsychological factors (see also Happé, 1993).

However, if one takes into account the ERP patterns elicited by sentence-initial scalars, a more complicated picture emerges. The exploratory analyses of ERP responses elicited by the sentence-initial scalar quantifiers suggest that pragmatic abilities influenced scalar statement processing already at the scalar quantifier. Perhaps counter-intuitively, differential processing of the two different scalar quantifiers was most pronounced in the pragmatically less skilled participants. We will provide more in-depth discussion of these issues in the general discussion, but what these results suggest is that pragmatically less skilled participants may have been able to temporarily ignore or inhibit their pragmatic knowledge during the processing of the critical words (see Feeney et al., 2004; Handley & Feeney, 2004), instead of being insensitive to pragmatic constraints (e.g., Schindeler et al., 2008).

In sum, our results suggest that pragmatic constraints can have rapid effects during on-line sentence comprehension. When pragmatic constraints are taken into account, as in low AQ-Comm people, they may guide expectations about upcoming words through the pragmatic presumption of informativeness. But when these constraints cannot be used or they are ignored, as in the high AQ-Comm group, the effects of other constraints may surface, such as the effect of lexical–semantic relationships. In our second experiment, we examined the incremental processing of weak scalar quantifiers further by modulating the effect of pragmatic constraints through linguistic focus.

## Experiment 2

### Introduction

Whereas blatantly underinformative statements that violate pragmatic principles are relatively uncommon in everyday language (perhaps with the notable exception of utterances where underinformativeness is used as a humoristic device), temporarily underinformative statements are quite common. For example, whereas a phrase such as “Some people have eyes,” is unlikely to appear, a sentence such as “Some people have eyes that are different colors” is much more natural.

In the sentence “Some people have eyes,” the comma signals clausal wrap-up and the end of the quantifier scope. This puts the clause-final words ‘eyes’ clearly into focus (e.g., Birch & Rayner, 1997; Hirotsani et al., 2006). In



contrast, in “Some people have eyes that are different colors”, the scope of the quantifier encompasses the whole relative clause construction (‘eyes that are different colors’) and the focus of the utterance – the part of the statement that the communicator wants to emphasize and is most relevant to the addressee for evaluating sentence meaning – is not ‘eyes’ but ‘different colors’.

Research on the role of focus in language comprehension suggests that the processing of unfocused materials is dominated by ‘low-level’ lexical–semantic relationships rather than by ‘full-fledged’ compositional processing that is needed to establish sentence truth-value or real-world plausibility (e.g., Ferreira, Bailey, & Ferraro, 2002; Sanford & Sturt, 2002). This is because readers and listeners generally devote less attention and processing time to unfocused material than to focused material (e.g., Cutler, Dahan, & van Donselaar, 1997; Frazier, Carlson, & Clifton, 2006), so that unfocused materials receive an incomplete semantic and pragmatic analysis (so-called shallow processing; e.g., Sanford & Garrod, 1998).

In Experiment 2, a second set of participants read sentences like “Some people have eyes that are different colors”. The first clauses of these sentences were identical to those used in Experiment 1, but the comma was excluded and the clause was always followed by a relative clause (see Table 2, for examples). Thus the sentences were informative overall but the first clause could be considered ‘locally’ underinformative. Given the absence of the comma and the fact that all scalar sentences in Experiment 2 had this same structure, we expected the critical words to be out of focus and we hypothesized that they would therefore be processed more shallowly (e.g., Sanford & Sturt, 2002), and the ERP response would be dominated by simple lexical–semantic relationships rather than by the pragmatic presumption of informativeness. In other words, we predicted that locally underinformative statements would fail to evoke a pragmatic N400 effect. Rather, we predicted that the N400 would be reduced, relative to the informative statements, because of their closer lexical–semantic relationships. In addition, we predicted that this effect would not be modulated by the real-world pragmatic abilities (AQ-Comm score) of the participants because the relevant pragmatic constraints were now the same in the informative and underinformative statements.

**Table 2**

Examples sentences from Experiment 2. Critical words are underlined for expository purpose only.

Underinformative/informative
Some people have <u>lungs/pets</u> that are diseased by viruses
Some rock bands have <u>musicians/groupies</u> with real drug problems
Some gangs have <u>members/initiations</u> that are really violent
Relatively poor/good semantic fit
Literature classes sometimes read <u>papers/poems</u> as a class
Wine and spirits contain <u>sugar/alcohol</u> in different amounts
Fillers
Many people catch the flu in the winter
Many vegetarians eat bean curd as a source of protein

## Methods

### Participants

Thirty-one right-handed Tufts students (13 males; mean age = 19.7 years) gave written informed consent. All were native English speakers, without neurological or psychiatric disorders, and had not participated in Experiment 1.

### Materials

We constructed 70 sentence pairs that were identical to the 70 critical sentence pairs from Experiment 1 up until and including the critical words (see Table 2). The 70 new sentences did not contain commas, and the critical words were always followed by a relative clause (e.g., “Some people have lungs that are diseased by viruses.”). In addition, we created 35 new filler sentences that, as in Experiment 1, started with the quantifier ‘many’ and involved a simple and true statement, and that, like the new ‘some’ sentences, did not contain a comma (e.g., “Many people catch the flu in the winter.”). To examine ERP responses to semantic fit, participants in Experiment 2 were also presented the exact same 70 sentences containing the semantic fit manipulation as used in Experiment 1.

### Procedure and EEG recording

The procedure of Experiment 2 was identical to that of Experiment 1 except for the presentation duration of the critical words (which was 227 ms longer in Experiment 1 due to the presence of commas).

EEG recording and pre-processing in Experiment 2 was identical to that in Experiment 1. Two participants were excluded due to excessive artifacts (mean trial loss >50%). For the remaining 29 participants, average ERPs (normalized by subtraction to a 100 ms pre-stimulus baseline) were computed over artifact-free trials for CWs in all conditions (mean trial loss across conditions 12%, range 0–35%, without substantial differences in mean trial loss across conditions).

The exit-interview in Experiment 2 was identical to that from Experiment 1 except for the last page. In Experiment 2, the last page gave subjects an example of a locally underinformative sentence and a locally informative sentence, and an explanation for why the first part of the sentence could be considered informative or underinformative. Subjects subsequently reported whether they had noticed *during the ERP experiment* that the first part of some sentences was odd for the above mentioned reason? If they answered ‘yes’ they were asked to report how consistently (on a 5-point scale) they noticed that some of these sentences sounded odd, whether they treated such underinformative sentences as true or false, and how consistently (on a 5-point scale) they treated these sentences as true or false.

## Results

### Main effect of informativeness

Critical words elicited larger N400 responses in the informative statements compared to the underinformative statements (see Fig. 4, left panel). Using mean amplitude across all electrodes in the 350–450 ms time window, a 2 (informativeness: informative, underinformative)  $\times$  2 (distribution: anterior, posterior) ANOVA revealed that informative statements elicited a larger N400 ERP than underinformative statements ( $F(1, 28) = 7.52, p = 0.011, CI - .56 \pm .42$ ), whereas this effect did not differ across anterior and posterior electrodes ( $F(1, 28) = .162, p = 0.690$ ). Separate ANOVAs for anterior and posterior electrodes, however, revealed that the main effect of condition was only marginally significant at anterior electrodes ( $F(1, 28) = 3.69, p = 0.065, CI -.53 \pm .56$ ) but fully statistically significant at posterior electrodes ( $F(1, 28) = 6.87, p = 0.014, CI -.65 \pm .51$ ).

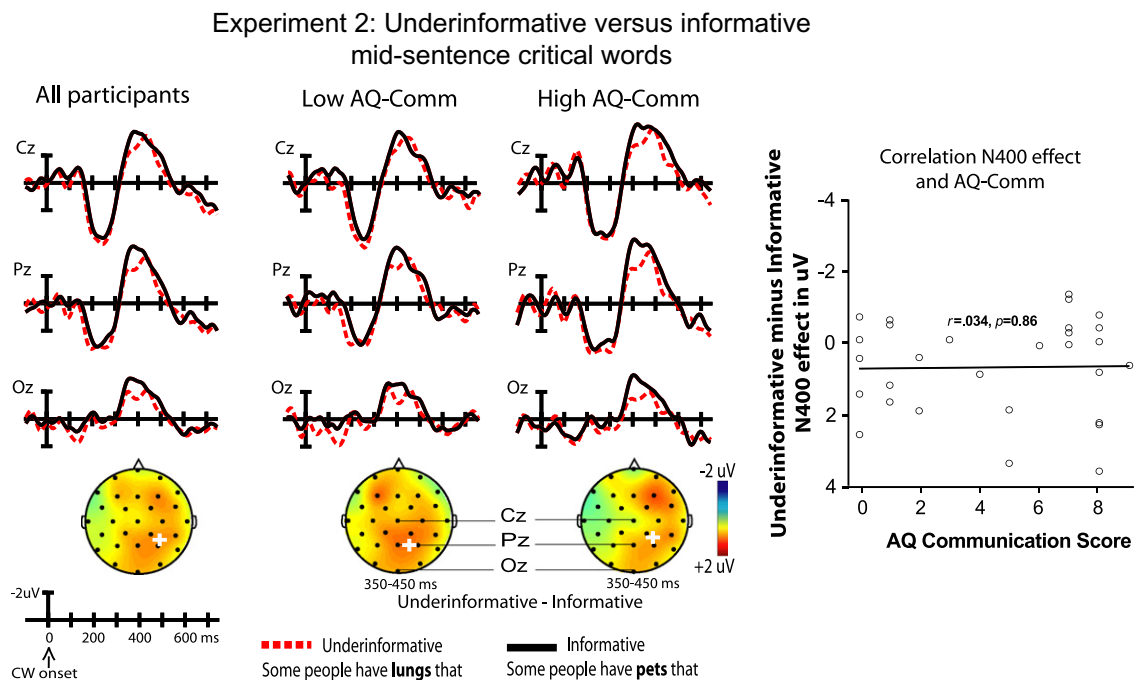
### AQ-Comm score and ERP responses to informativeness

AQ scores ranged from 5 to 36 ( $M = 21.17, SD = 7.88$ ). The participants were again grouped into low AQ-Comm ( $N = 14$ ) and high AQ-Comm ( $N = 15$ ) groups based on the median split of scores on the Communication subscale. AQ-Comm score for the low AQ-Comm group ranged from 0 to 5 ( $M = 1.43, SD = .43$ ; five males and nine females, mean age 20.4 years, mean total AQ score 15.9), and from 6 to 9 for the high AQ-Comm group ( $M = 7.4, SD = .25$ ; 7

males and 8 females, mean age 19.2 years, mean total AQ score 26.1). The two AQ-Comm groups showed statistically significant differences in AQ-Comm score ( $t(27) = 12.18, p < .001$ ) and in total AQ score ( $t(27) = 4.49, p < .001$ ), as well as in age ( $t(27) = 2.43, p < .05$ ). As in Experiment 1, the factor age, when entered into the subsequent analyses as a covariate, did not change the patterns of results.

Grand average ERPs for the two groups are displayed in Fig. 4 (right panel). Using mean amplitude at posterior electrodes in the 350–450 ms time window, the overall 2 (informativeness: informative, underinformative)  $\times$  2 (AQ-Comm Group: low AQ-Comm, high AQ-Comm) ANOVA revealed no significant interaction effect ( $F(1, 27) = .712, p = .41$ ), suggesting that the groups similarly showed larger N400 responses to informative compared to underinformative statements. Consistent with this result, and in contrast to Experiment 1, pragmatic abilities now did *not* predict the size of the underinformative N400 effect, as there was no significant correlation between AQ-Comm score and the mean ERP difference score for underinformative and informative statements (Pearson's  $r = .034, p = .86$ ), nor between the ERP difference score and scores on the total AQ score ( $r = -.11, p = .57$ ) or any of the AQ subscales (Social Skill,  $r = -.09, p = .63$ ; Attention-Switching,  $r = -.13, p = .516$ ; Imagination,  $r = 0.78, p = .69$ ; Attention To Detail,  $r = -.282, p = .139$ ).

To directly test for differential effects of AQ-Comm group across the two experiments, we performed a 2 (informativeness: informative, underinformative)  $\times$  2 (AQ Group: low AQ, high AQ)  $\times$  2 (Experiment: Experiment 1, Experiment 2) ANOVA. This analysis revealed a statistically



**Fig. 4.** Left panel: Grand average ERPs elicited by critical words in underinformative (dotted lines) and informative (solid lines) statements from Experiment 2. Right panel: Grand average ERPs elicited by critical words in underinformative and informative statements per AQ-Communication group in Experiment 2.

significant 3-way interaction effect ( $F(1, 54) = 8.29$ ,  $p = .006$ ), supporting the observation that AQ-Comm modulated the effect of informativeness in Experiment 1 but not in Experiment 2.

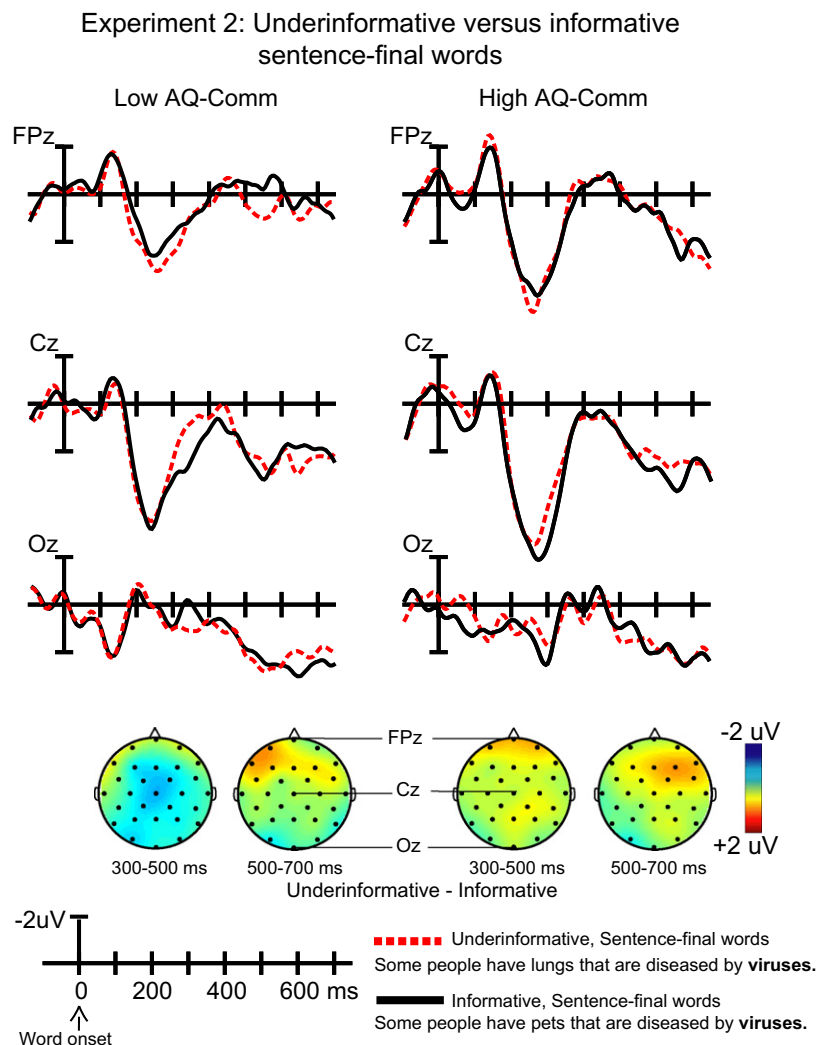
#### ERP responses to informativeness and the role of LSA

We repeated the same analyses as in Experiment 1 to investigate the role of lexical-semantic factors, and computed ERP responses for one item set that had a relatively small LSA difference between informative and underinformative sentences, and one set that had a relatively large LSA difference between conditions (see <http://www.nmr.mgh.harvard.edu/kuperberglab/materials.htm>). The results indicated that across both groups, LSA modulated the effect of informativeness ( $F(1, 28) = 5.31$ ,  $p = .029$ ) such that there was an effect of informativeness in the item set with large LSA differences between conditions ( $F(1, 28) = 9.35$ ,  $p = .005$ ) but not in the item set with small

LSA differences between conditions ( $F(1, 28) = .31$ ,  $p = .58$ ). These results did not differ between the two groups ( $F(1, 27) = .88$ ,  $p = .36$ ).

#### Group differences in ERP responses to sentence-final words

We also examined the ERP responses to sentence-final words in underinformative and informative statements between the two AQ-Comm groups. As shown in Fig. 5, there was no modulation of ERPs evoked by sentence-final words in the underinformative relative to the informative statements in either the low or the high AQ-Comm group. Using mean amplitude in the 300–500 ms time window across all electrodes, the overall ANOVA revealed no significant main effect of informativeness ( $F(1, 28) = .13$ ,  $p = .72$ ) and no significant interaction effect of informativeness with AQ-Comm group ( $F(1, 27) = .92$ ,  $p = .35$ ). Repeating the above analyses for sentence-final words across the 500–700 ms window yielded the same pattern of results.



**Fig. 5.** Grand average ERPs elicited by sentence-final words in underinformative (dotted lines) and informative (solid lines) statements per AQ-Communication group in Experiment 2, shown at electrode locations FPz, Cz, and Oz.

### Group differences in ERP responses to real-world congruity

As in Experiment 1, the two AQ-Comm groups produced similar real-world congruity N400 effects (see Fig. 6). Using mean amplitude at posterior electrodes in the 350–450 ms time window, the overall 2 (Real-world congruity: congruous, incongruous)  $\times$  2 (AQ-Comm Group: low AQ-Comm, high AQ-Comm) ANOVA revealed a significant main N400 effect of real-world congruity ( $F(1, 28) = 23.05, p < .001, CI -1.66 \pm .73$ ) but no significant congruity by group interaction effect ( $F(1, 27) = 2.26, p = .15$ ). Also, as in Experiment 1, there was no significant correlation between AQ-Comm score and the size of the real-world congruity N400 effect ( $r = .15, p = .44$ ).

In a direct test for differential effects of AQ-Comm group across the two experiments, a 2 (Real-world congruity: congruous, incongruous)  $\times$  2 (AQ Group: low AQ, high AQ)  $\times$  2 (Experiment: Experiment 1, Experiment 2) ANOVA did not reveal a significant 3-way interaction effect ( $F(1, 54) = 1.41, p = .24$ ), i.e. the two AQ-Comm subgroups showed the same effects of real-world congruity in the two experiments.

### Exploratory analyses for ERP responses to the scalar quantifiers

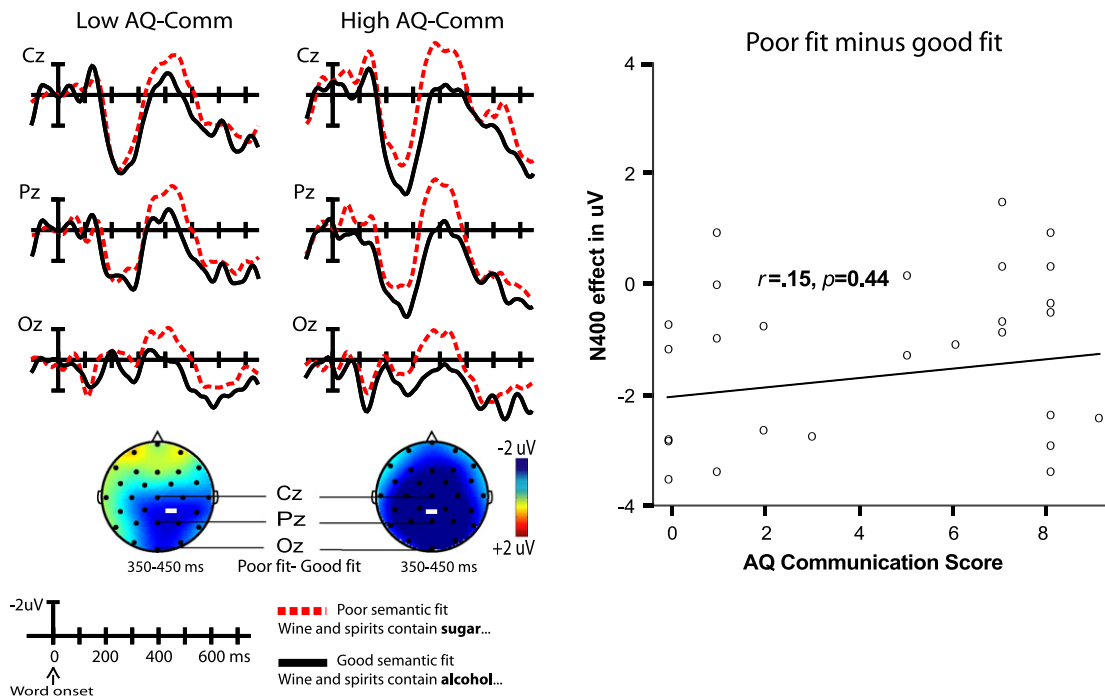
As in Experiment 1, we examined whether the two groups differed in their ERP responses to the sentence-initial quantifiers (the results are available at <http://www.nmr.mgh.harvard.edu/kuperberglab/materials.htm>).

The quantifier ‘Some’ elicited a relatively broadly distributed negativity compared to ‘Many’ in both groups. We performed a series of repeated measures ANOVAs to test for a 2 (quantifier: some, many) by 2 (AQ-Comm group: low AQ-Comm, high AQ-Comm) interaction, in adjoining 50 ms time windows between 100 and 800 ms after quantifier onset, using all electrodes or only anterior or posterior electrodes. Only in the 450–500 ms window was there a marginally significant interaction effect when using all electrodes ( $F(1, 27) = 4.2, p = .053$ ), or only anterior electrodes ( $F(1, 27) = 2.99, p = .095$ ). Follow-ups showed that the quantifier ‘some’ elicited a more negative ERP compared to ‘many’ in the low AQ-Comm group (all electrodes,  $F(1, 13) = 4.314, p = .058, CI -.85 \pm .87$ ; anterior electrodes,  $F(1, 13) = 5.353, p = .038, CI -1.07 \pm 1.00$ ), but not in the high AQ-Comm group (all electrodes,  $F(1, 14) = .564, p = .465, CI .29 \pm .27$ ; anterior electrodes,  $F(1, 14) = .017, p = .90, CI -.05 \pm .81$ ). Additional analyses showed that there was a marginally significant correlation between AQ-Comm and the differential quantifier effect when using all electrodes ( $r = -.366, p = .051$ ). The differential ERP effect at the quantifier further predicted the differential ERP effect at the critical word when using posterior electrodes ( $r = -.435, p = .018$ ), but also the differential ERP effect of real-world congruity ( $r = -.51, p = .005$ ).

### Exit-interview

We examined whether the AQ-Comm groups differed in their exit-interview ratings for truth-value and natural-

### Experiment 2: Poor semantic fit versus good semantic fit



**Fig. 6.** Left panel: Grand average ERPs elicited by words that had a relatively poor (dotted lines) and relatively good (solid lines) semantic fit per AQ-Communication group in Experiment 2 and corresponding scalp distributions. Right panel: Correlation between N400 effect and AQ-Communication score.



ness. As in Experiment 1, the 2 (AQ-Comm group: low AQ-Comm, high AQ-Comm) by 2 (informativeness: informative, underinformative) ANOVA revealed no group differences in the truth-value ratings and the naturalness ratings (all  $F_s < 2$ ). In addition, underinformative and informative statements received similar truth-value ratings ( $t < 1$ ) but different naturalness ratings ( $t(1, 28) = 9.3$ ,  $p < .001$ ).

## Discussion

In Experiment 2, critical words in statements that were temporarily underinformative but out of discourse focus elicited a smaller N400 than critical words in informative statements. This effect was not modulated by the pragmatic language abilities of the participants, but was modulated by the lexical–semantic differences between conditions. We take these results to suggest that, when statements were out of focus (in this case due to the sentence structure and scalar quantifier scope), initial semantic processing costs were driven primarily by the lexical–semantic relationships between each critical word and the previous words in the sentence, rather than pragmatic constraints of informativeness. Interestingly, whereas all participants in Experiment 1 had indicated in the exit-interview to have noticed underinformativeness, none of the participants in Experiment 2 indicated to have noticed any ‘local’ underinformativeness and there were no differential processing costs on sentence-final words.

As mentioned in the introduction to Experiment 2, we think that the fact that all scalar sentences in Experiment 2 had the same relative clause construction may have contributed to the relatively shallow processing of critical words, as compared to Experiment 1. We designed the experiment in this way so that participants would expect all scalar sentences to have a particular structure with the most important information near the end of the sentence, directing their focus away from the critical words. These experiment-based expectations may be related to structural priming in comprehension whereby the syntactic structure of a sentence can influence the analysis of subsequent sentences (e.g., Branigan (2007), for review).

As in Experiment 1, we found an interaction effect between pragmatic abilities and the ERP response to the sentence-initial quantifiers in Experiment 2, suggesting that the groups differed in their pragmatic response to the quantifiers. However, this was not a clear-cut replication. Whereas the high AQ-Comm group showed a larger differential effect of quantifier type in Experiment 1, it was the low AQ-Comm group that showed a larger differential effect of quantifier type in Experiment 2. In addition, although the differential effects of quantifier type were most pronounced at anterior electrodes in both experiments, the statistically significant interaction effects occurred in different time windows across the two experiments. Of course, the experiments differed in an important way, namely that in Experiment 1 the quantifier ‘some’ was associated with potential underinformativeness downstream, but not in Experiment 2. It is therefore possible that the different patterns of results across the experi-

ments reflect that certain task-strategies that were relevant in Experiment 1 (e.g., pragmatic processing related to the differences in informativeness between ‘Some’ and ‘Many’) were not applicable in Experiment 2.

Taken together, the results from our first and second experiment suggest that contextual factors, whether these are derived from individual pragmatic abilities or the overall experimental context (see also Breheny et al., 2006), and lexical–semantic factors modulate the processing of scalar statements. Moreover, when contextual factors attenuate the impact of pragmatic underinformativeness, either because certain participants are less likely to process ‘pragmatically’ or because the experimental context makes local underinformativeness go unnoticed, lexical–semantic factors are more likely to surface.

## General discussion

In Experiment 1, we tested the hypothesis that pragmatically underinformative statements incur a semantic processing cost as indexed by the N400 ERP component, and, moreover, that this is more likely to happen in healthy individuals who are relatively pragmatically skilled than in healthy individuals who report everyday-life pragmatic communication difficulties. Across all participants, underinformative and informative statements (e.g., “Some people have lungs/pets, ...”) elicited similar N400 ERPs, but this absence of an overall effect was due to opposite effects in participants depending on their pragmatic abilities. Pragmatically more skilled participants (the low AQ-Comm group) showed a larger N400 to underinformative versus informative statements – a pragmatic N400 effect that was independent of the lexical–semantic differences between the underinformative and informative conditions. In contrast, pragmatically less skilled participants (the high AQ-Comm group) showed a larger N400 for informative versus underinformative statements, and this effect was driven by lexical–semantic factors because it disappeared when we controlled for lexical–semantic differences between conditions. Interestingly, in Experiment 1, processing differences between these two groups were already observable at the scalar quantifiers (differential effects on ‘some’ versus ‘many’ in the high AQ-Comm but not the low AQ-Comm group), and were also evident at the end of the sentences (a reduced effect on the sentence-final word in the high AQ-Comm relative to the low AQ-Comm group). In contrast, the two groups showed no differences in their N400 response to statements with a relatively poor versus good semantic fit in relation to real-world knowledge (e.g., “Wine and spirits contain sugar/alcohol ...”).

In Experiment 2, we examined the role of linguistic focus in pragmatic processing by comparing ERP responses to the same underinformative statements followed by a relative clause construction (e.g., “Some people have lungs/pets that...”). Because of the larger quantifier scope in these sentences, the local underinformativeness of the embedded statement was irrelevant to ongoing processing and, according to our exit-interview, went unnoticed by all participants. As expected, the informative statements now elicited larger N400 ERPs than (locally) underinformative

statements, and this N400 modulation was strongly dependent on lexical–semantic factors in pragmatically more and less skilled participants alike.

### Implications for theory

The fact that underinformative statements elicited an N400 effect compared to informative statements, albeit only in a subgroup of participants, suggests that the pragmatic meaning of scalar quantifiers ('some but not all') can rapidly and incrementally contribute to sentence comprehension (see also Grodner et al., 2010), at least to the extent that the pragmatic meaning was available when the critical was encountered. This pragmatic N400 effect is inconsistent with an early claim in the experimental pragmatic literature that pragmatic scalar meaning results from a post-semantic decision process (e.g., Noveck & Posada, 2003), and with theoretical accounts of language processing that assume an initial, purely linguistic-semantic analysis that is followed by a later pragmatic stage of processing (e.g. Fodor, 1983; Forster, 1979).

The implications of our results for Relevance Theory or for the Levinsonian account, however, are less clear as neither theory makes explicit predictions about the time course of inferential processes. In a Levinsonian account, the pragmatic scalar meaning is thought to be automatically generated without having to traverse through a logical interpretation first (e.g., Levinson, 2000), whereas Relevance Theory assumes that scalar inferences are made only when they are sufficiently supported by the discourse context (e.g., Sperber & Wilson, 1986). It has been argued by some researchers that because single sentences like "Some people have lungs" are without a discourse context (i.e., 'neutral'), and because the logical interpretation is the default because of its simplicity. Thus, one interpretation of Relevance Theory might predict an initial logical interpretation and a delay in pragmatic interpretation (e.g., Breheny et al., 2006). The fact that we observed a pragmatic N400 effect on 'lungs' in the low AQ-Comm Group in Experiment 1 suggests that there was no such initial logical interpretation at the point of the critical word. It suggests that, at least in these participants, the pragmatic meaning of 'some' was immediately available at the point of the critical word in the absence of a discourse context.

On the other hand, there were several aspects of the data that are consistent with an interpretation of Relevance Theory that emphasizes the roles context, standard of relevance and that allows for certain anticipatory processes. First, the experiments themselves may be considered as 'global context' which, in principle, could have biased some participants towards making scalar inferences more frequently and rapidly than usual in Experiment 1 and discouraged these participants to generate such inferences in Experiment 2. Second, advocates of Relevance Theory have argued that the presumption of relevance may guide certain anticipatory processes (e.g., Wilson & Sperber, 2004), and that pragmatic scalar meaning can be generated without having to traverse through a logical interpretation first (see also Noveck & Sperber, 2007). Third, what is relevant or not will depend on the specific reader or listener and can therefore differ across individu-

als and groups. Each of these points is discussed in further detail below.

### Effects of experimental context

In both Experiments 1 and 2, many scalar sentences were presented in close succession. In Experiment 1, one might argue that the presentation of so many underinformative sentences biased the low-AQ participants towards generating more scalar inferences than they would during normal language comprehension. This, however, seems unlikely: there is little support from the experimental literature that participants in experimental settings are biased towards generating scalar inferences. In fact, in studies using sentence-verification tasks, at least 40–50% of participants do not generate scalar inferences (e.g., Bott & Noveck, 2004; Noveck & Posada, 2003), perhaps because the true/false verification encourages a strategic use of formal logic (cf. Feeney et al., 2004). Also, our exit-interview in Experiment 1 indicated that *all* participants mentioned that they had treated the underinformative scalar statements as being true. This, if anything, suggests that the experimental context may have biased *against* the generation of scalar inferences.

In Experiment 2, however, the experimental context biased against the generation of scalar inferences limited to the critical word only, but not against the generation of scalar inferences per se. The experimental context in this experiment consisted of the repeated presentation of scalar statements with a quantifier scope that extended beyond the critical word, and therefore establishing truth-value of the complete proposition was only possible at a later moment in time. The result of this attenuation of the impact of local pragmatic underinformativeness by contextual factors was that the local underinformativeness went unnoticed, and that lexical–semantic factors dominated semantic processing of the critical words.

### Incremental processing and pragmatic expectancy

The observed relationship between pragmatic abilities and the N400 ERP was specific to pragmatic underinformativeness, because we found no relationship between pragmatic abilities and the N400 modulation by real-world semantic fit or modulation of the N400 in Experiment 2. These results suggest that the reported N400 modulation of informativeness is due of the 'genuinely pragmatic' violation of Gricean maxims (Grice, 1975) instead of a violation of or deviation from real-world semantic knowledge. Thus, in addition to other factors, the N400 indexes the lexico-semantic processing consequences of using pragmatic constraints during on-line sentence comprehension.

In line with recent work, we have suggested that the pragmatic presumption of informativeness guides the participant's expectations about upcoming words (e.g., Nieuwland & Kuperberg, 2008; Van Berkum, 2009; Van Berkum et al., 2005), which facilitates subsequent semantic processing. Such expectations may be generated before the onset of the critical word, i.e. they may constitute active predictions that facilitate lexical access of the critical word (see Delong et al., 2005; Van Berkum et al., 2005),

or the relevant information may be retrieved only once the critical word is presented (e.g., Brown, Hagoort, & Kutas, 2000; for discussion on these two different interpretations of the N400, see Federmeier & Kutas, 1999; Kutas & Federmeier, 2000; Kutas et al., 2006; Lau et al., 2008; Van Berkum, 2009). Although much of the prediction literature deals with lexically specific predictions during language comprehension (e.g., DeLong et al., 2005; Van Berkum et al., 2005), we take our results as potential evidence for relatively coarse-grained anticipation, a background of expectations of relevance that can be revised or elaborated as sentences unfold (e.g., Wilson & Sperber, 2004).

We propose that, in the low-AQ-Comm participants, the availability of pragmatic scalar meaning allowed participants to derive expectations about upcoming words based on the pragmatic presumption of informativeness, so that they expected the upcoming word to denote something that only some, hence not all people have. In this respect, we take our N400 results in these subjects in Experiment 1 not to directly reflect full-fledged, online pragmatic inferencing, but rather to reflect the semantic processing consequences of earlier and relatively implicit pragmatic inferencing (see also Kuperberg, Choi, Cohn, Paczynski, & Jackendoff, 2010a; Kuperberg, Paczynski, et al., 2010b; Van Berkum, 2009).

In addition to the effects at the critical word, there were also additional downstream processing consequences of violating pragmatic expectations, with effects on the sentence-final words in both low AQ-Comm and high AQ-Comm groups (although, as discussed below, this effect was somewhat smaller in the high AQ-Comm group). No such effect was seen in Experiment 2. Our design did not allow us to test specific hypotheses with regard to these sentence-final effects and the nature of this differential ERP effect remains unclear. We think it is unlikely to be an N400 effect, because of its more frontal distribution and prolonged morphology. Instead, it could reflect a larger positivity to sentence-final words in the underinformative statements. This would be consistent with other reports of positive ERP effects elicited by sentence-final words in sentences requiring inferencing as compared simpler sentences (e.g., Filik, Sanford, & Leuthold, 2008; Kuperberg, Choi, et al., 2010a; Kuperberg, Paczynski, et al., 2010b), possibly reflecting additional sentence wrap-up processing (e.g., Steinhauer & Friederici, 2001).

#### *Individual differences in scalar processing*

One of the most striking findings of Experiment 1 was the heterogeneity between the individuals in their ERP profiles that was predicted by their real-life pragmatic abilities. One possible interpretation of these results is that scalar inferences are not obligatory (see also Bott & Noveck, 2004; Breheny et al., 2006): many of the participants in Experiment 1 – the less pragmatically skilled participants – did not show a pragmatic N400 effect. This result seems to mirror the observation from the behavioral literature that some people tend to make scalar inferences and some do not (e.g., Noveck & Posada, 2003), as defined by a

'false' or a 'true' response to underinformative scalar statements.

The underlying cause of these individual differences is as yet unknown and such variability in the healthy adult population has often been ignored by researchers (but see Banga, Heutinck, Berends, & Hendriks, 2009; Feeny et al., 2004). One relatively trivial explanation for these individual differences is that the groups differed in the extent that they were actually paying attention to sentence meaning rather than to the superficial coherence or lexical-semantic relatedness of the words in the sentences. This seems unlikely, however, for several reasons. First, such a 'non-specific attention' account would predict similar patterns for Experiments 1 and 2. In Experiment 2, however, we saw no between-group differences even though the same lexical items were presented. Second, such an account would predict group differences in the ERP responses to the real-world congruity manipulation. Even though the real-world congruous–incongruous sentences differed in several respects from the scalar sentences, one might expect some differential N400 effects between the two groups for these items, given that the N400 is sensitive to attentional factors (see Chwilla, Brown, & Hagoort, 1995). Third, as discussed below, the high AQ-Comm participants showed differential effects at the point of the quantifier itself in Experiment 1, suggesting that they were attending to its meaning.

We therefore suggest that a more specific impairment mediated the absence of a pragmatic N400 effect in the high AQ-Comm participants of Experiment 1. One possibility is that these participants were unable to generate scalar inferences, e.g., Noveck, 2001). This draws analogies from research on the differential ability of children and adults to generate scalar inferences. Young children (age 7–9 years) seem less likely than older children or adults to generate scalar inferences (e.g., Guasti et al., 2005; Noveck, 2001; Pouscoulous, Noveck, Politzer, & Bastide, 2007; Smith, 1980), although it appears that this ability is largely constrained by task-specific features and that it can be improved by training (e.g., Feeny et al., 2004; Guasti et al., 2005; Pouscoulous et al., 2007). Such results are often ascribed to children having a relatively under-developed ability for pragmatic inferencing or a relative insensitivity to pragmatic constraints (e.g., Smith, 1980). If the high AQ-Comm participants in the current study show similar impairments, these may, in turn, be related to the notion of 'standard of relevance' from Relevance Theory. This holds that individuals generate scalar inferences only when they are required to meet the individual's internal standard of relevance, reflecting a trade-off between the possible cognitive gains associated with generating the inference and the amount of cognitive effort necessary to derive it (e.g., Carston, 1998). Individuals with self-reported impaired pragmatic abilities may have a lower standard of relevance, to the extent that they are less likely to compute the pragmatic consequences of linguistic input. It could also be the case that generating scalar inferences is more costly for those individuals, as has been suggested for children (e.g., Noveck, 2001).

However, our ERP findings indicate that the high AQ-Comm group was not completely insensitive to the

pragmatic manipulation. Both AQ-Comm groups showed a sentence-final ERP modulation by informativeness (although this effect was marginally larger in low AQ-Comm participants). In addition, both low and high AQ-Comm participants indicated in the exit-interview that they had in fact registered the pragmatic anomaly. There are at least two different explanations for these results. First, it is possible that the high AQ-Comm participants simply showed a delay in pragmatic processing. For example, they might not have generated a scalar inference by the time the critical word was encountered (leading to a pragmatic N400 effect) but such an inference may have been computed at some later point such that, at the sentence-final word, the pragmatic anomaly was registered. This account, however, is difficult to reconcile with the observation that these high AQ-Comm participants did show a differential ERP effect at the scalar quantifier itself.

An alternative explanation is that, the high AQ-Comm participants did register the pragmatic meaning of the word 'some', but strategically ignored or inhibited the resulting pragmatic meaning of the scalar statements at the point of the critical word, and focused on the logical meaning of the sentences instead, perhaps based on their observation that standard conversational norms in Experiment 1 were repeatedly violated (see Guasti et al., for a similar suggestion). This latter explanation is similar to what has been proposed to explain individual differences in logical reasoning tasks, namely that some people are simply better in temporarily ignoring or inhibiting their pragmatic knowledge in order to focus on the logical reasoning requirements of a task (see Feeney et al., 2004; Handley & Feeney, 2004; Stanovich & West, 2000). The fact that the high, but not low, AQ-Comm group showed a differential ERP response to the scalar quantifiers could be taken as consistent with such an account. These early ERP effects may have reflected the active 'undoing' of the automatic access to the pragmatic meaning of 'some'.

These conclusions, however, are only speculative, and dedicated follow-up experiments are needed to examine the functional significance of the observed ERP differences the scalar quantifiers. One possible prediction is that high AQ-Comm people are also more likely to respond 'true' to underinformative statements in a sentence-verification paradigm (but see Pijnacker et al., 2009) and that this behavioral outcome is heralded by ERP responses to the scalar quantifiers.

Although our experiments were not optimized for investigating this issue and our conclusions are necessarily post hoc, the ERP modulations at the scalar quantifiers might be taken to reflect pragmatic processing that took place before encountering the critical words. Although the nature of the differential effect of quantifier differed between Experiments 1 and 2, in both experiments there was a near-significant correlation between the differential effect of quantifier and the differential effect at the critical words. This correlation should be interpreted with caution, however, because it could be the case that larger overall ERP amplitudes also generate larger differential effects, which may confound the examination of a relationship between two ERP difference scores. Nevertheless, one tentative conclusion could be that scalar quantifiers can rapidly evoke scalar inferences

that guide expectations about upcoming words based on the pragmatic presumption of informativeness, and, as a result, influence downstream semantic processes.

#### *A dynamic interplay between levels of processing, as indexed by the N400*

In addition to the pragmatic effect of (under)informativeness, this paper has also highlighted two other influences on semantic processes as indexed by the N400: the effects of lexico-semantic co-occurrence, and the effect of real-world knowledge. Previous studies have shown that all these factors can independently influence the modulation of the N400, and that, when they all support the same interpretation, they can act in parallel to facilitate processing (e.g., Ditman, Holcomb, & Kuperberg, 2007b; Federmeier & Kutas, 1999). A key question, however, is which of these factors prevail when they are in conflict with one another. There are reports of lexical-semantic associations overriding any effect of pragmatic constraints on the N400 (e.g., Fischler et al., 1983; Kounios & Holcomb, 1992; Noveck & Posada, 2003), but this seems to be the case only in pragmatically infelicitous sentences (see Nieuwland & Kuperberg, 2008, for discussion). Lexico-semantic associations can, under other circumstances, also temporarily dominate processing of implausible sentences and discourse, with delayed effects observable within a late positivity/P600 time window (e.g., Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Nieuwland & Van Berkum, 2005; for review see Kuperberg (2007)). However, in a recent study, we showed that causal coherence across sentences can modulate the N400, even when semantic relationships between individual words are matched (Kuperberg, Choi, et al., 2010a; Kuperberg, Paczynski, et al., 2010b). The evidence thus points towards a dynamic interplay between different levels of processing, with each level of processing being influenced by a range of relevant factors. In the current study, we have shown that pragmatic licensing can override lexical-semantic co-occurrence in some individuals but not in others. Moreover, we have shown that differences in linguistic focus can shift the balance from 'full-fledged', higher-order compositional processing to processing driven by lexical-semantic relationships (e.g., Ferreira et al., 2002; Sanford & Garrod, 1998; Sanford & Sturt, 2002). Taken together, these observations provide further evidence for a dynamic interplay between lexical-semantic, pragmatic and neuropsychological factors during on-line sentence comprehension.

#### **Conclusion**

A major feat of human cognition is our ability to use language to efficiently communicate about the world. Mapping an incoming message about the world onto our world knowledge involves at least two aspects: the message can be true or false with respect to what we hold to be true, and it can be relatively informative or trivial in the light of what we already know. In the case of scalar statements, this means that logical-structural meaning of the scalar quantifier needs to be combined with our real-



world knowledge, our pragmatic knowledge of what constitutes a trivial or informative thing to say, and our individual tendencies to rely more on logic or pragmatic aspects of language. In the current study, we provide ERP evidence that all these factors exert their influence during on-line language comprehension.

## Acknowledgments

We are very grateful to Kana Okano, Liam Clegg and Sarah Cleary for their help with stimulus preparation and data collection, and to Ira Noveck, two anonymous reviewers and Andrea Eyleen Martin-Nieuwland for helpful comments on earlier drafts of this manuscript. This research was supported by an NWO Rubicon Grant to MSN and by NIMH-R01-MH071635 to GRK.

## References

- Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorder*. Washington, DC: American Psychiatric Association.
- Bach, K. (2006). The top 10 misconceptions about implicature. In B. J. Birner & G. Ward (Eds.), *A festschrift for Larry Horn*. Amsterdam: John Benjamins.
- Banga, A., Heutinck, I., Berends, S. M., & Hendriks, P. (2009). Some implicatures reveal semantic differences. In B. Botma & J. van Kampen (Eds.), *Linguistics in the Netherlands* (pp. 1–13). Amsterdam: John Benjamins.
- Baron-Cohen, S. (2008). Autism, hypersystemizing, and truth. *Quarterly Journal of Experimental Psychology*, *61*, 64–75.
- Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (2000). *Understanding other minds: Perspectives from autism and developmental cognitive neuroscience*. Oxford: Oxford University Press.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism spectrum quotient (AQ): Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17.
- Bezuidenhout, A., & Cutting, J. C. (2002). Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, *34*, 433–456.
- Bezuidenhout, A., & Morris, R. (2004). Implicature, relevance and default inferences. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics*. Basingstoke: Palgrave Macmillan.
- Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *28*, 653–660.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*, 55–73.
- Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437–457.
- Branigan, H. P. (2007). Syntactic priming. *Language and Linguistics Compass*, *1*, 1–16.
- Breheny, R., Katsos, N., & Williams, J. (2006). Are generalized scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*, *100*, 434–463.
- Brown, C. M., Hagoort, P., & Kutas, M. (2000). Postlexical integration processes in language comprehension: Evidence from brain-imaging research. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 881–895). Cambridge, Mass.: MIT Press.
- Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research*, *1146*, 172–184.
- Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Amsterdam: John Benjamins.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond*. Oxford: Oxford University Press.
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*, 274–285.
- Clark, H. H. (1996). *Using language*. New York, NY, US: Cambridge University Press.
- Coulson, S. (2004). Electrophysiology and pragmatic language comprehension. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 187–206). San Diego: Palgrave Macmillan.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. M. N. Zwicky (Eds.), *Natural language parsing* (pp. 320–358). Cambridge, UK: Cambridge Univ. Press.
- Cutler, A., & Clifton, C. E. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford: Oxford University Press.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*, 141–201.
- De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology*, *54*, 128–133.
- 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.
- Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007a). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology*, *44*, 927–935.
- Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007b). The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors. *Language and Cognitive Processes*, *22*, 793–827.
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language*, *54*, 554–573.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491–505.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469–495.
- Feeeny, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology*, *58*, 121–132.
- Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Filik, R., Sanford, A. J., & Leuthold, H. (2008). Processing pronouns without antecedents: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*, 1315–1326.
- Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*, 400–409.
- Fischler, I., Childers, D. G., Acharyyapaopan, T., & Perry, N. W. (1985). Brain potentials during sentence verification: Automatic aspects of comprehension. *Psychophysiology*, *20*, 400–409.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum.
- Francis, W., & Kucera, H. (1982). *Frequency analysis of English usage*. New York: Houghton Mifflin.
- Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*, 244–249.
- Gamut, L. (1991). *Logic, Language, and Meaning* (Vol. 1). Chicago, Illinois: University of Chicago Press.
- Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.
- Geurts, B. (2009). Scalar implicature and local pragmatics. *Mind and language*, *24*, 51–79.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Grodner, D., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). “Some”, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*, 42–55.
- Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*, 667–696.

- Hagoort, P., Hald, L. A., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, 1146, 210–218.
- Handley, S. J., & Feeney, A. (2004). Reasoning and pragmatics: The case of even if. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics*. London: Palgrave Press.
- Happé, F. G. E. (1993). Communicative competence and theory of mind in autism. A test of relevance theory. *Cognition*, 48, 101–109.
- Hirotoni, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54, 425–443.
- Horn, L. R. (1972). *On the semantic properties of logical operators in English*. Ph.D. thesis, University of California, Los Angeles.
- Horn, L. R. (2006). The border wars: A neo-Gricean perspective. In K. von Stechow & K. Turner (Eds.), *Where semantics meets pragmatics* (pp. 21–48). Amsterdam: Elsevier.
- Huang, Y., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, 58, 376–415.
- Joliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: Linguistic processing in high-functioning adults with autism or Asperger syndrome: Is local coherence impaired? *Cognition*, 71, 149–185.
- Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54, 445–465.
- 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.
- Kucera, J., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49.
- Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., & Jackendoff, R. (2010a). Electrophysiological correlates of Complement Coercion. *Journal of Cognitive Neuroscience*, doi:10.1162/jocn.2009.211333.
- Kuperberg, G. R., Paczynski, M., & Ditman, T. (2010b). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, doi:10.1162/jocn.2010.21452.
- Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 17, 117–129.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463–470.
- 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.
- Kutas, M., Van Petten, C., & Klueder, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York: Elsevier.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Landauer, T. K., Foltz, P. W., & Dumais, S. T. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284.
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the n400. *Nature Reviews Neuroscience*, 9, 920–933.
- Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading—XVI. The visual span in normal and low vision. *Vision Research*, 37, 1999–2010.
- Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lord, C., Risi, S., Lambrecht, L., et al. (2000). The autism diagnostic observation schedule – Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview – Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 659–685.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Musolino, J., & Lidz, J. (2006). Why children are not universally successful with quantification. *Linguistics*, 44, 817–852.
- Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19, 1213–1218.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary change deafness in discourse comprehension. *Cognitive Brain Research*, 24, 691–701.
- Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78, 165–188.
- Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language*, 85, 203–210.
- Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, 12, 425–431.
- Noveck, I. A., & Sperber, D. (2004). *Experimental pragmatics*. Basingstoke: Palgrave.
- Noveck, I. A., & Sperber, D. (2007). The why and how of experimental pragmatics: The case of 'scalar inferences'. In N. Roberts (Ed.), *Advances in pragmatics*. Palgrave.
- Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, 1153, 166–177.
- Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based lexical anticipation: Prediction or priming? *Discourse Processes*, 45, 464–496.
- Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, 39, 607–618.
- Pouscolous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). Processing costs and implicature development. *Language Acquisition*, 14, 347–375.
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology*, 53A, 1061–1080.
- Recanati, F. (2003). Embedded implicatures. *Philosophical Perspectives*, 17, 299–332.
- Regel, S., Gunter, T. C., & Friederici, A. D. (2010). Isn't it ironic? An electrophysiological exploration of figurative language comprehension. *Journal of Cognitive Neuroscience*. doi: 10.1162/jocn.2010.21411.
- Rips, L. J. (1975). Quantification and semantic memory. *Cognitive Psychology*, 7, 307–340.
- Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26, 159–190.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, 6, 382–386.
- Schindele, R., Lüdtke, J., & Kaup, B. (2008). Comprehending negation: A study with adults diagnosed with high functioning autism or Asperger's syndrome. *Intercultural Pragmatics*, 5, 421–444.
- Sedivy, J. (2007). Implicatures in real-time conversation: A view from language processing research. *Philosophy Compass*, 2, 475–496. doi: 10.1111/j.1747-9991.2007.00082.x.
- Smith, C. L. (1980). Quantifiers and question answering in young children. *Journal of Experimental Child Psychology*, 30, 191–205.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Basil Blackwell.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645–726.
- Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 30, 267–295.
- Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders*, 11, 45–56.

- Tager-Flusberg, H. (1985). Psycholinguistic approaches to language and communication in autism. In E. Schopler & G. B. Mesibov (Eds.), *Current issues in autism* (pp. 69–88). New York: Kluwer.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). San Diego, CA, US: Academic Press, Inc.
- Van Berkum, J. J. A. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things? In M. Carreiras, Jr. & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229–270). New York: Psychology Press.
- Van Berkum, J. J. A. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 31, 443–467.
- Van Berkum, J. J. A., Van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580–591.
- Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, 8, 485–531.
- Van Petten, C., Weckerly, J., Mclsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, 8, 238–242.
- Wilson, D., & Sperber, D. (2004). Relevance theory. In G. Ward & L. Horn (Eds.), *Handbook of pragmatics* (pp. 607–632). Oxford: Blackwell.
- Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*, 35, 331–335.