

What can electrophysiology tell us about the cognitive processing of scalar implicatures?

Abstract

One of the most widely studied phenomena in neuropragmatics—the study of how the brain derives context- and speaker-based aspects of meaning—is *scalar implicature*. A scalar implicature is the interpretation of a proposition like *Some of the students failed* as meaning that a stronger proposition (*Not all of the students failed*) is not true. While scalar implicatures have been a significant object of study for decades, in recent years in particular there has been an explosion of experiments investigating them using neuroscientific methods, particularly electroencephalography (EEG). Much of this research aims to identify neural substrates of comprehending scalar implicatures. Here I review the extant findings and argue that most of these studies have not directly observed neural correlates of scalar implicatures; rather, they have mostly observed downstream and/or domain-general processes that happen to be related to implicatures, but not uniquely so. I argue that an instrumental approach to neuroscience—one that treats brain components not as objects of research in of themselves, but as tools for learning about pragmatics—would be a valuable addition to this emerging field.

Keywords: scalar implicatures, experimental pragmatics, neurolinguistics, event-related potentials, ERPs, N400, cognitive neuroscience

Introduction

The past decade has seen an explosion of research using neuroscience methods to investigate how people process semantic and pragmatic aspects of language. A challenge that arises in such a research enterprise is that both the object of study (language, and the psychological mechanisms that handle it) and the tools (patterns of brain activity) are not fully understood, and the link between them is even less understood. These challenges raise potential concerns about how much can be learned about the processing of language using neuroscience techniques. These concerns are not new – for example, Poeppel and Embick (2005) discuss fears that the alliance of linguistics and neuroscience could result in unproductive "interdisciplinary cross-sterilization". The present paper revisits this concern by adopting as a case study one particular subfield of neurolinguistics that has attracted substantial interest in the last decade: the neural processing of *scalar implicature*.

Scalar implicature is a phenomenon whereby a given utterance is interpreted as implying that some stronger utterance is not true (Horn, 1972; Levinson, 2000).

For example, if a speaker says *Some of my students failed*, a hearer often infers that the speaker must have meant that not all of their students failed; this inference is putatively based on the fact that the speaker chose not to say *All of my students failed* even though that would have provided more specific information if it were true. Substantial experimental research has been devoted to determining the nature of the computations underlying this inference—i.e., whether this inference is made by default or only in certain contexts, whether the inference-based meaning is realized automatically and effortlessly or at a processing cost, and whether this computation happens immediately or at a delay (see, e.g., Chemla & Singh, 2014a,b; Katsos & Cummins, 2010; Noveck & Reboul, 2008; Sauerland & Schumacher, 2016).

Scalar implicature has long been one of the most studied phenomena in experimental pragmatics – indeed, the study of scalar implicature played a seminal role in the emergence of "experimental pragmatics" as a field. Much research on the computation of scalar implicatures relies on behavioural measures, including explicit judgments about the meanings of various utterances under various conditions (Bott & Noveck, 2004; Bott, Bailey, & Grodner, 2012; Chemla & Spector, 2011; Chevallier et al., 2008; De Neys & Schaeken, 2007; Degen, 2015; Dieussaert, Verkerk, Gillard, & Schaeken, 2011; Doran, Baker, McNabb, Larson, & Ward, 2009; Feeney, Scafton, Duckworth, & Handley, 2004; Geurts & Pouscoulous, 2009; Goodman & Stuhlmüller, 2013; Marty & Chemla, 2013; Marty, Chemla, & Spector; van Tiel, van Miltenburg, Zevakhina, & Geurts, 2016; among others), reaction and reading time measures (Bergen & Grodner, 2012; Bezuidenhout & Cutting, 2002; Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Chemla, Cummins, & Singh, 2017; Hartshorne & Snedeker, ms.; Noveck & Posada, 2003; Politzer-Ahles & Fiorentino, 2013; Politzer-Ahles & Husband, in press; among others), measures of eye movements (Breheny, Ferguson, & Katsos, 2012, 2013; Degen & Tanenhaus, 2014; Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang & Snedeker, 2009; among others), and mouse tracking (Tomlinson, Bailey, & Bott, 2013). Recent reviews of these literatures are available in, e.g., Chemla & Singh (2014a,b) and Sauerland & Schumacher (2016). In recent years there has also been explosion of pragmatics research using neuroscience methods, particularly event-related brain potentials (see Table 1).

Event-related potentials¹ are a powerful tool for studying language processing, and particularly the comprehension of scalar implicatures, for several reasons (for an introduction to event-related potentials in language comprehension,

¹ As well as their magnetic equivalent, event-related fields. Event-related potentials are electric signals recorded using electroencephalography, whereas event-related fields are magnetic signals recorded using magnetoencephalography. While there are some differences in the techniques, especially in terms of what parts of the brain these techniques are most capable of measuring signals from, the techniques are conceptually very similar and have similar applications. In the following discussion I consider both of these together—sometimes under the umbrella term *electrophysiology*, which typically includes both electroencephalography and magnetoencephalography, and which distinguishes these methods from hemodynamic methods (such as functional magnetic resonance imaging) which have substantially different properties.

see Kaan, 2007). They have excellent temporal resolution, allowing researchers to distinguish between brain activity that occurs at the moment a scalar implicature is made and activity that occurs later during downstream processing. They are multidimensional since signals may differ in terms of where on the scalp they appear, how late after a stimulus they emerge, how long they last, whether they have positive or negative voltage relative to the recording reference, and how strong they are; this provides a means to distinguish between qualitatively different processes. And they can be recorded without any task that requires participants to make metalinguistic judgments about the meaning of sentences, which is especially important given that sentences with different interpretations (e.g., where *some* is interpreted with or without a scalar inference) may engender different downstream verification strategies during judgment tasks—for instance, determining whether *some and possibly all Xs are Y* requires just searching through the Xs until the comprehender finds one that meets the criterion Y, whereas determining whether *some but not all Xs are Y* requires searching for one that does meet the criterion Y and one that does not (see, e.g., Politzer-Ahles & Gwilliams, 2015).

With the power of neuroscientific methods, however, also come limitations.² The past decade or so of EEG research on scalar implicatures serves as an instructional example of some of the shortcomings and pitfalls that can occur when attempting to use neuroscience methods to address linguistic or psychological questions. These shortcomings are not just abstract; they have real consequences for the potential value of neurolinguistic research on pragmatics. For example, some active researchers in the experimental pragmatics field have lamented that the copious outpouring of experimental studies on scalar implicature through the years has actually had little impact on theories of how scalar implicature works. As the popularity of using neuroscience techniques to test linguistic and psychological questions is unlikely to go away in the foreseeable future, there is a real need to find ways to make neuro-pragmatics experiment provide more useful and impactful observations for shaping our understanding of how pragmatics, the mind, and the brain work.

The present paper will summarize what has been observed so far in studies using these methods, as well as caveats about what these studies cannot show. While this review focuses on scalar implicatures as a case study, the issues discussed here can apply to pretty much any situation in which a linguistic or psycholinguistic phenomenon is investigated using electrophysiological methods. For anyone beginning to learn about any given topic in neurolinguistics, or preparing to do experiments themselves on such topics, it is valuable to have an awareness of the potential pitfalls of this kind of research and familiarity with other ways this kind of research can be done.

² For similar challenges in the application of hemodynamic neuroscientific methods to research on other kinds of inferences, see Virtue & Sundermeier (2016).

Scalar implicatures: linguistic and psycholinguistic questions

Before looking into the ways neuroscience methods have been used to study scalar implicatures, it will be informative to consider what some of the core questions about scalar implicatures are. Scalar implicatures have been the subject of substantial debate regarding both their linguistic properties and the psycholinguistic mechanisms that support their comprehension. For more detailed reviews, the reader is referred to Chemla & Singh, 2014a,b; Katsos & Cummins, 2010; Sauerland, 2012; and Sauerland & Schumacher, 2016; among others.

One major question is what kind of implicature a scalar implicature really is. An implicature is an interpretation of an utterance which arises not because of the actual meanings of the words in the utterances and the ways in which they are combined, but because of something else. That "something else" may be one of many different things, and thus there are many different kinds of implicature. Consider, for instance, a vignette like the following:

"Carlyle asked Thisbe if she would lend him a thousand dollars. Thisbe scoffed and walked out of the room."

The second sentence not only expresses the literal meaning of the two clauses (Thisbe scoffed, Thisbe walked out of the room), but also implies that (1) Thisbe walked out of the room after scoffing, rather than before, and (2) Thisbe is not willing to lend Carlyle a thousand dollars. According to the theory introduced by Grice (1989), these two implicatures arise via different mechanisms. The implicature that Thisbe scoffed before leaving the room, as opposed to after, arises from a linguistic convention, called *conjunction buttressing*, whereby two clauses conjoined with "and" are often interpreted as describing events occurring one after the other (e.g., Levinson, 2000:122). Grice (1989) calls this a "generalized conversational implicature" because it is a function of the use of particular words and phrases themselves, rather than a function of a specific context, and thus can arise in most contexts. That is to say, almost any time someone says "*X scoffed and walked out of the room*", it will be interpreted as meaning X scoffed first and walked out of the room later, unless something else about the context of that utterance makes this interpretation unlikely or impossible. On the other hand, the implicature that Thisbe is not willing to lend Carlyle a thousand dollars depends on this particular context, and would not have arisen if "*Thisbe scoffed and walked out of the room*" had been uttered in most other contexts. Grice (1989) calls this a "particularized conversational implicature". Some theories consider scalar implicatures, such as the implicature that *some of my students failed* means *not all of my students failed*, to be a type of generalized conversational implicature, which would mean that they are realized via different mechanisms than many other types of pragmatic meaning. Other theories, however, either propose that scalar implicatures are not generalized conversational implicatures, or eschew the distinction between generalized and particularized implicatures entirely. In either case, such approaches predict that scalar implicatures are derived by the same

context-specific mechanisms as any other implicatures (see, e.g., Degen & Tanenhaus, 2015; Noveck & Sperber, 2007).

While the question of whether or not scalar implicatures are generalized conversational implicatures mainly concerns linguistic representation rather than psychological processing, this issue has also raised a closely related question in psycholinguistics: that of how quickly and effortfully scalar implicatures are understood in real-life comprehension. Much of this research has assumed that if scalar implicatures are a kind of generalized conversational implicature then it follows that they would be realized immediately and effortlessly by comprehenders, whereas if they are dependent on particular contexts then it will take extra time and extra processing costs for a scalar implicature to arise since the comprehender must first evaluate the context (see, e.g., Breheny et al., 2006; Katsos & Cummins, 2010). A large portion of the behavioural research on scalar implicatures has been aiming to resolve this question, using a variety of psycholinguistic techniques to test whether scalar implicatures arise earlier or later than literal meanings, and whether they evoke extra processing costs. More recent formulations have complicated this question by pointing out that even if scalar implicatures are context-dependent they may not necessarily always take extra time or processing resources (Degen & Tanenhaus, 2015), and that the derivation of a scalar inference is not a monolithic process, but is a process involving multiple intermediate steps, each of which might be rapid and effortful or might be slow and costly (Chemla & Singh, 2014a,b). Some recent research is beginning to make steps towards evaluating these more articulated kinds of scalar implicature processing models; for example, Rees and Bott (2018) use the psycholinguistic paradigm of *structural priming* to investigate the properties of a particular sub-component of the scalar implicature derivation process, the determination of which alternatives are relevant in the context (e.g., whether or not *all* is a relevant alternative to *some* in the discourse).

Another major question about scalar implicatures is whether they are derived by pragmatic or semantic mechanisms. All the approaches mentioned above assume that scalar implicatures are derived pragmatically, i.e., by integrating information about the context with information from the linguistic expressions themselves to infer what the speaker meant. But other approaches hold that scalar implicatures are derived not by making context-based inferences about what the speaker might have meant, but from language itself. One way this might happen is through the insertion of a covert semantic operator into the sentence structure (Chierchia, Fox, & Spector, 2012). Another is by the selection of a lexicalized enriched meaning—for example, for the interpretation of *some* as *some but not all*, it could be the case that *some* is polysemous, with one sense equivalent to some-and-possibly-all, and another sense equivalent to some-but-not-all, and a "scalar implicature" is just the selection of the latter sense (see e.g. Sauerland, 2012).

Under such accounts, a so-called "scalar implicature" is not an implicature at all.³ Another major line of research in the experimental pragmatics of scalar implicatures, then, focuses on testing whether people derive scalar inferences in certain contexts, such as embedded clauses, where pragmatic and semantic accounts make different predictions about whether or not scalar implicatures should be derived (for review see Chemla & Singh, 2014a,b; Sauerland, 2012; among others).

These questions about the linguistic representation and psychological processing of scalar implicatures provide a backdrop for the recent explosion of neurolinguistics research into the phenomenon.

What previous event-related potential studies have shown

Since the seminal study by Noveck and Posada (2003), and particularly in the last decade, many studies have been conducted attempting to use neuroscientific data to adjudicate between competing theories of how scalar implicatures are processed, and to identify the neural correlates of scalar implicature realization. Table 1 summarizes all the studies, to my knowledge, that have used electrophysiology to study scalar implicature processing. (While the present review focuses on electrophysiology, the points that will be raised below regarding the functional interpretation of brain data apply likewise to data from hemodynamic studies using, e.g., functional magnetic resonance imaging.)

Insert Table 1 here

The incremental use of scalar inferences to drive downstream predictions

Many studies of scalar implicature have focused on brain responses to words occurring later in an utterance which serve as probes of whether or not an enriched meaning (e.g., *some but not all*) has been realized earlier. In Table 1, I refer to this paradigm as "Violation of SI-based prediction". While such studies do not attempt to directly identify the brain correlate of the inference-making process itself, this sort of downstream probe of whether or not an inference was made is useful for testing theories of when and how scalar implicatures occur. For example, Noveck and Posada (2003) compared event-related potentials elicited by the final words of patently true sentences like *Some people have brothers* (it is true that some, but not all, people have brothers) and those of infelicitous sentences like *Some staircases have steps* (while the sentence is literally true, it is not true that "not all

³ In this paper I will continue to refer to these enriched meanings, such as "*some but not all*", as "implicatures" or "inferences" when discussing research that tests when and how such meanings are realized. This sort of shorthand is conventional in much of this literature.

staircases have steps"). Later studies introduced more controlled designs (e.g., Nieuwland et al., 2010), with the most recent studies in this vein using picture stimuli to control the contexts rather than relying on world knowledge as a context like the above examples do (Hunt et al., 2013; Spychalska et al., 2016, 2019). For example, Hunt and colleagues (2013) showed participants visual contexts in which an agent affected none of one group of referents (i.e., the brownies in the lower example), all of another group (i.e., the tomatoes in the lower example), and some-but-not-all of the third group (i.e., the steaks in the lower example); see Figure 1. Participants then read a sentence like "The boy cut some of the steaks in this story". The critical word could felicitously refer to a group that the boy cut some but not all of (the lower example), or infelicitously refer to a group that the boy actually cut all of (the upper example). The logic of this and other such studies is that if a scalar implicature has been made, such that *some* is interpreted as *not all*, then the readers should be surprised to read the infelicitous critical word (e.g., *steaks* in the upper example in Figure 1), given that they would expect the sentence to mention the group that the boy cut some but not all of (e.g., *brownies* in the upper example in Figure 1). If the scalar implicature is realized as the sentence is unfolding, this surprise should elicit a different brain response as soon as the critical word is read. On the other hand, if scalar implicatures are only realized after a proposition has been fully uttered, then such an effect might not be observed when the critical word is read.

 Insert Figure 1 here

In the majority of these studies, downstream words that render a sentence infelicitous (*steaks* in the upper example in Figure 1) indeed elicit a different brain response than downstream words that render a sentence true (*steaks* in the lower example). Typically, they tend to elicit a stronger N400 response. The N400 is an event-related potential component observed around 300-500 ms after the appearance of a contentful stimulus, mostly on central and posterior parts of the scalp, and tends to be stronger when a stimulus is less expected based on the context (for review see Lau, Phillips, & Poeppel, 2008, and van Berkum, 2009, among others). Thus, these studies have shown that the inference-based interpretation of *some* is integrated incrementally into the online interpretation of a sentence, modulating the extent to which downstream words are expected. Words consistent with both the literal interpretation and the enriched "not all" interpretation are more strongly expected than words consistent only with the literal, not the enriched, interpretation; the more expected words, in turn, evoke smaller N400 components than the less expected words that render a sentence underinformative. These studies have provided important evidence that scalar implicatures are realized before the end of a sentence. This finding goes against the

predictions of strong pragmatic accounts which assume that implicatures are a function of a whole utterance, although it can still be accommodated in modern pragmatic accounts (Geurts, 2010). This technique is also useful as a probe for how scalar implicatures influence comprehension in different populations; for instance, Spychalska and colleagues (2016) found that comprehenders who tend to interpret *some* as *not all* in an explicit judgment task also show a different neural response profile in this paradigm than participants who tend not to use this interpretation in the explicit judgment task.

Converging evidence for the incremental use of scalar inferences to update processing predictions comes from Hartshorne and colleagues (2014), who adapted a paradigm widely used in reading time research (Bergen & Grodner, 2012; Breheny et al., 2006; Hartshorne & Snedeker, ms.; Politzer-Ahles & Fiorentino, 2013; Politzer-Ahles & Husband, in press) to show that the derivation of scalar implicatures is sensitive to context. In this paradigm, people read a sentence in which *some* is embedded in a context that either does or does not support scalar implicatures, and then a later word in the sentence serves as a probe for whether the inference was realized. For example, in the sentence *Addison ate some of the cookies before breakfast this morning, and the rest are on the counter*, "some" is likely to be interpreted as meaning "not all", and thus the reader should be aware that there are some cookies that have not been eaten. The comprehension of *the rest* is then facilitated since the reader was already aware of this remaining set of referents. On the other hand, in the sentence *If Addison ate some of the cookies before breakfast this morning, then the rest are on the counter*, "some" is less likely to be interpreted as meaning "not all", because scalar inferences are less likely in downward entailing contexts such as the antecedent of a conditional (i.e., an "if..." statement; see Chierchia et al., 2012, among others). Thus, in this context, comprehension of *the rest* is not facilitated. Indeed, the authors observed a more negative electrophysiological signal at *the rest* in the latter case, and not in maximally similar control comparisons that do not involve scalar implicature. This is consistent with the notion that the scalar implicature was realized in the former context and facilitated later processing of *the rest*, whereas it was not realized in the latter case and thus *the rest* was less expected and more difficult to process.

Differences between pragmatic and semantic processing

While the abovementioned studies probed the downstream consequences of scalar implicatures rather than the neural correlates of the scalar implicatures themselves, other studies have attempted to directly measure brain responses underlying the realization of scalar implicatures. These too have the potential to test different theories of scalar implicature representation and processing: if one or more neural correlates of scalar implicature processing could be identified, then the properties of these responses could tell us something about the nature of scalar implicature processing. One such body of work has used event-

related potentials to examine whether scalar implicatures elicit qualitatively different processing than semantics. Most of this work has done so by comparing scalar implicature-based violations to purely semantic violations. In Table 1, I refer to this paradigm as "Infelicitous scalar expression". For example, Politzer-Ahles, Fiorentino, Jiang, and Zhou (2013) examined brain responses to putatively pragmatic violations, along with matched control conditions for each of these, at the first moment the scalar implicature could have been realized, rather than at a downstream word; see Figure 2. Pragmatic violations were realized by showing the participant a context in which everybody is doing the same thing (i.e., all of the girls are sitting on blankets), and then presenting them with a sentence like "*In this picture, some of the...*". If *some of the* is interpreted as meaning *not all of the*, then this sentence should be considered infelicitous by the comprehender, and should elicit a different brain response than a correct control (the same sentence if preceded by a picture in which some, but not all, of the girls are sitting on blankets). Indeed, these pragmatic violations elicited sustained negative brain responses compared to the correct controls. Semantic violations, on the other hand, were realized by showing the participant a context in which different people are doing different things (some girls are sitting on blankets and some are not) and then presenting them with a sentence like "*In this picture, all of the...*". Unlike the pragmatic violations, these sentences should be considered incorrect under any circumstance, given that the meaning of *all* is not negotiable like the meaning of *some* is. These violations did not elicit the sustained negativity, relative to their correct controls, that the pragmatic violations elicited. Crucially, these effects were observed at the quantifier itself, where the scalar implicature might first become available, rather than downstream as in the studies summarized above.

Such findings have been taken as evidence that pragmatic and semantic processing engender qualitatively different brain responses, a claim which has also drawn support from hemodynamic brain imaging research that has found activation in different brain regions for putatively pragmatic vs. semantic violations in a similar paradigm (Shetreet, Chierchia, & Gaab, 2014a, b, c; Zhan, Jiang, Politzer-Ahles, & Zhou, 2017). Such a dissociation would be theoretically important, given that it is an open question whether scalar implicatures are derived by pragmatic and semantic mechanisms, as discussed above. Isolating the neural mechanisms that are involved in scalar implicatures but not in other sorts of computations could provide a means to determine whether those mechanisms are involved in pragmatic or semantic processes, by comparing the neural correlates of scalar implicature processing to the neural correlates of other processes that are uncontroversially pragmatic or uncontroversially semantic.⁴ However, the event-related potential pattern observed in this study was not replicated in a similar study by Panizza, Onea, and Mani (2014), and only in a subset of participants in Politzer-Ahles (2013, Experiment 3). Furthermore, as discussed further below, it remains unclear whether

⁴ Such an enterprise would, however, rely on what are known as reverse inferences (Poldrack, 2006), which are of questionable validity.

the presence of qualitatively different processing components for "pragmatic" and "semantic" violations reflects qualitatively different correlates of pragmatic and semantic processing, or different domain-general processes that are differentially implicated in these two types of violations.

Insert Figure 2 here

How scalar implicatures are actually realized

Finally, some studies have compared the brain responses elicited by a scalar expression like *some* in contexts that do or do not license scalar inferences, in order to isolate components involved in the derivation of scalar inferences themselves as opposed to the deployment of these interpretations for downstream prediction or the processing of pragmatically infelicitous stimuli. While the studies described above tested scalar implicatures indirectly by using violation paradigms where the interpretation based on a scalar implicature conflicts with some other aspect of the context or world knowledge, these studies attempt to directly probe what happens when scalar implicatures actually occur. The motivation for such studies is similar to that for the studies described in the previous section: if we can identify the neural substrates of the scalar implicature derivation process itself, we could be able to use that to learn more about the nature of that process. In Table 1, I refer to this paradigm as "Contextual support for inference".

Various context manipulations have been used to carry out this sort of test. As described above, Hartshorne and colleagues (2014) examined brain responses to *some* in sentences like "*Addison ate some of the cookies...*" and "*If Addison ate some of the cookies...*". It has been frequently observed that scalar implicatures are typically not realized in an "if" clause, like in the latter example (Chierchia et al., 2012, among others). Therefore, if deriving a scalar implicature requires extra processing effort, *some* in the former context might elicit an additional neural response not elicited by *some* in the latter context. This is not, however, what has been observed. Hartshorne and colleagues (2014) found no significant difference in brain responses to *some* in contexts that do or do not license scalar inferences. Politzer-Ahles & Gwilliams (2015), using a similar experimental paradigm but a different context manipulation, found a sustained magnetoencephalogram component, originating from left lateral prefrontal cortex, for *some* in contexts that are less likely to license scalar inferences, compared to contexts that are more likely—the opposite of the prediction outlined above. They argued that this component reflects increased activation related to deriving the "not all" interpretation of *some* when it has little contextual support. Further research is

needed, however, both to determine the replicability of this component and to investigate whether this reflects pragmatic processing specifically, or domain-general operations that simply happen to also be implicated in a scalar implicature manipulation.

Are these "pragmatic components"?

A limitation of many of these studies, however, is that the brain responses observed are not uniquely attributable to pragmatics, as opposed to being domain-general processes; this is also acknowledged by Hartshorne and colleagues (2014), Hunt and colleagues (2013), and Nieuwland and colleagues (2010). Consider, for example, studies that found modulation of the N400 on critical words downstream of a scalar expression. Hunt and colleagues (2013), for instance, observed that the word *steaks* in the sentence "*The boy cut some of the steaks in this story*" elicited a larger N400 in a context where all of the steaks were cut (and thus the words *steaks* is not expected here) than in a context where some, but not all, of the steaks were cut (and thus the word *steaks* highly expected); see Figure 1. The N400, however, is not a direct index of scalar implicature processing. Rather, it is known to be an index of the ease or difficulty of lexical activation.⁵ Thus, these experiments are not directly observing brain responses related to pragmatics. Rather, they are observing domain-general brain responses which happen to be influenced by pragmatic processes that have already happened. That is to say, the realization of a scalar inference makes a given word less expected and less congruent with the context, and therefore harder to access; it is the difficulty of access, not the scalar implicature itself, that leads to a greater N400. Fortunately, those studies were not designed to directly observe scalar implicature processing itself, and the conclusions of those studies do not depend on being able to observe scalar implicature processing itself. Readers of this literature must be careful not to misinterpret these studies as revealing neural correlates of scalar implicatures themselves.

We could think of the brain responses in such experiments as being like a high-tech thermometer. A person who is sick might show a higher reading on a thermometer than a person who is not sick. But this happens because the sickness causes the person's body temperature to rise, and that temperature in turn affects the thermometer. A person observing the thermometer reading is not directly observing the illness; rather, they are only observing a downstream consequence of it. In the same way, a person observing N400 effects in an experiment is observing

⁵ This is a simplification; for more nuanced review see, e.g., Lau, Phillips, and Poeppel (2008) and van Berkum (2009), among others. In fact the N400 is not even specific to language; it can be elicited by nonlinguistic stimuli such as pictures (Ganis & Kutas, 2003; Goto, Ando, Huang, Yee, & Lewis, 2010) and smells (Kowalewski & Murphy, 2012).

downstream consequences of a pragmatic computation, rather than observing the computation itself.

Another relevant comparison is eye-movement research on scalar implicatures (e.g., Grodner et al., 2010; Huang & Snedeker, 2009). In these studies, a camera monitors how participants' eyes move as they are viewing a display, under the assumption that when a comprehender realizes a certain interpretation they will look towards the image depicting that interpretation. When participants move their eyes to look at a given picture in a visual-world display, this presumably reflects the consequence of a pragmatic computation: making a scalar inference allows the hearer to decide who is being referred to in the sentence, and later they look at that referent. Moving their eyes does not reflect the pragmatic computation itself. In eye movement research, however, this is widely understood, and eye movements are generally not mistaken for representing pragmatic processing itself. In neurolinguistic research, on the other hand, perhaps because brain components are less well understood, there is a temptation to interpret brain components as directly reflecting pragmatic processing—and indeed much research aims to identify the brain regions and brain components that carry out scalar implicatures. As described above, though, brain components observed in such experiments usually cannot be uniquely attributed to pragmatic processing itself.

The problem is more serious with experiments observing new or unpredicted components. Studies examining N400 effects on downstream critical words, such as those discussed above, are designed to elicit specific downstream effects (i.e., N400s) and thus are in a good position to interpret these effects. For those studies, the fact that the brain responses do not reflect pragmatic processing is not a problem; those studies were designed to teach us something about pragmatic processing without needing to directly observe "pragmatic" brain responses. Other studies, have aimed to directly observe what brain responses are elicited by pragmatic processing. The latter two categories of studies described above are both subject to this concern. Politzer-Ahles and colleagues (2013), for example, compared brain responses elicited by pragmatic vs. semantic violations, with no specific predictions about how they would differ. They then observed qualitatively different patterns for the two violation types, and concluded that there are qualitatively different neural mechanisms for the processing of semantics and pragmatics; indeed, this study is frequently cited as offering experimental support for that view. But do these responses really show that pragmatics and semantics are processed in different ways? An alternative explanation is that these pragmatic and semantic violations elicited different *domain-general* processing strategies. For example, the pragmatic violation included a sentence that could be re-interpreted to fit the context: the context was a picture in which everyone is doing the same thing (e.g., there are five girls, and all of them are sitting on blankets) and the critical sentence was *In the picture, some of the girls are sitting on blankets*; see Figure 2. If *some of* is interpreted as meaning "at least one", rather than as meaning "some but not all", then the sentence is no longer inconsistent with the

context. Thus, the different brain response elicited in this violation might have reflected the reinterpretation of the sentence, or the inhibition of one of the interpretations. If that is the case, then this brain response does not reflect pragmatics per se; rather, what has been attributed to "pragmatic processing" is rather an epiphenomenon of other domain-general processes, such as revision, that happened to be possible in this kind of violation sentence and not in the semantic violation sentence. Crucially, in a different kind of manipulation it could turn out that semantic violations and not pragmatic violations could implicate this process.⁶ Similar alternative, domain-general explanations are available for (and suggested in) other electrophysiological studies that attempted to probe scalar implicature processing directly (e.g., Hartshorne et al., 2014; Politzer-Ahles & Gwilliams, 2015). Such studies provide an empirical observation (e.g., that two types of violation are processed differently) but do not provide an explanation for that observation. Unfortunately, results from such studies have sometimes been claimed to provide evidence for what the neural correlates of scalar implicatures are. In fact, as described above, such a conclusion is not strictly justified.

These concerns have been raised before (e.g., van Berkum, 2009, 2010), and in fact are just a special case of more general problems raised by Poeppel and Embick (2005): the *granularity mismatch problem* and the *ontological incommensurability problem*. The granularity mismatch problem refers to the limitations of searching the brain for coarse-grained concepts like "pragmatics" and "semantics" in a language system that works with more fine-grained, specific operations. Much of the neurolinguistics work in the past two decades has been resolving this problem by focusing on more and more specific phenomena—rather than searching for neural correlates of syntax and semantics, nowadays we have research subfields searching for the neural correlates of gender agreement, complement coercion, scalar implicature, etc., just to name a few. There is, however, a second problem, the ontological incommensurability problem: the brain likely functions in terms of basic operations like linearization, concatenation, chunking, etc., rather than abstract linguistic operations like syntactic movement, implicature generation, etc. Thus, such research may be examining units that are incommensurable, looking for correlates of things like scalar implicature in a brain that does not have such an operation.

An alternative approach

There is another approach to neurolinguistic research on scalar implicatures, which could be more fruitful. This approach is, essentially, not to look for "neural

⁶ For example, semantic scope ambiguities (e.g., whether "every kid saw a painting" means that different kids saw different paintings or they all saw the same painting) might involve similar processes of revision or adjudication between multiple interpretations; while it is not clear to me what the maximally similar "pragmatic" control comparison would be, such semantic ambiguities have been found to elicit sustained negativities downstream (Dwivedi et al., 2010).

correlates" of scalar implicatures, but to use brain components as a thermometer: something which we recognize is not a correlate of scalar implicatures, but which has well-understood properties that allow it to be used as an independent tool for testing the predictions of theories.

To put this approach in context, a categorization of existing approaches is useful. van Berkum (2010) classifies neuro-pragmatics research into four rough categories. "Neuro lite" research involves doing psycholinguistic research (with behavioural rather than neuroscientific measures) but just phrasing the conclusions as being about the "brain" rather than about the "mind". "Instrumental" research involves using neuroscientific measures not to understand the brain per se, but as a tool to understand psychological processes. "Modestly ontological" research involves searching for neural correlates of some process or concept that is motivated by linguistic theory. Finally, "deeply ontological" research attempts to understand how actual brain functions support these processes. Much extant neurolinguistics research on scalar implicatures, especially the studies attempting to observe scalar implicature processing directly, could be considered "modestly ontological", as it often takes the brain responses in of themselves as the object of study; this type of research is subject to the limitations described above. On the other hand, an "instrumental" approach would be valuable for integrating neuroscientific methods with pragmatic theories and psychological models of scalar implicature processing.

How does an instrumental approach to neuroscience work in practice? An excellent example in neurolinguistics comes from van Turennout and colleagues (1998), who used event-related potentials to detect when people accessed different components of a lexical entry. Specifically, they wanted to see whether a person comprehending a word accesses its morphosyntax first or its pronunciation first. They did so by taking advantage of the lateralized readiness potential, a brain response that emerges when a person prepares to make a muscle movement, such as moving the hand to press a button. This component emerges even if the person ultimately decides not to make the movement; thus, it can be used to reveal things that a person considers doing but ultimately does not do. In this study, Dutch native speakers saw a picture representing some word (e.g., a picture of a bear), and their task was to press a button to indicate the word's morphological gender. Unless, that is, the first sound in the word was /b/, in which case participants were supposed to do nothing. A lateralized readiness potential was observed even on trials with no button press, suggesting that participants prepared to press a button to indicate gender, even though they ultimately did not press the button. In other words, participants accessed the morphosyntactic gender of the word, and accordingly prepared to make a response, before they accessed the pronunciation of the word and stopped preparing their response. Crucially, the lateralized readiness potential itself has nothing to do with the processing of morphosyntax or phonology; it is a simple correlate of motor preparation and its properties are well known. Nonetheless, the experimenters were able to design a paradigm in which

this component could be used as a simple instrument to test much more abstract aspects of cognition.

In fact, much electrophysiological research on scalar implicatures is already using this approach. The body of research examining how the realization of a scalar inference modulates processing of words downstream (Hunt et al., 2013; Nieuwland et al., 2010; Noveck & Posada, 2003; Sikos et al., 2013; Spsychalska et al., 2016, 2019), summarized above, exemplifies an instrumental approach: such studies were never intended to show what is happening in the brain when a scalar implicature is actually realized, but rather were intended to use downstream brain responses as an instrument to detect whether or not an inference had been realized earlier. The interpretation of these results is, accordingly, more straightforward than the interpretation of much of the rest of the electrophysiological literature discussed above. This paradigm has convincingly demonstrated that scalar implicatures can be realized before the end of a proposition, and has been fruitfully used to examine how scalar implicatures are or are not derived in different populations (e.g., Spsychalska, 2016). However, examining downstream responses can only answer some, not all, of the important questions about scalar implicatures. By now, most theories of scalar implicature processing agree that *some* can be realized as *not all* without the comprehender needing to wait until the end of the proposition; even strongly pragmatic accounts have apparatus to account for such phenomena (e.g., Geurts, 2010). In current debates, the issue at stake regarding the speed of scalar implicatures is not so much a question of whether they are realized at the end of a sentence or in the middle, but whether they are realized immediately or a few hundred milliseconds later (see, e.g., Huang & Snedeker, 2009). Such a question cannot be answered by looking at brain responses a few words after a scalar implicature could have been triggered; it requires paradigms that look directly at the expression that triggers the inference. Likewise, examining downstream modulations of how strongly a word is expected might not be sufficient for determining what sort of computation a scalar implicature itself is (e.g., semantic or pragmatic). For reasons like this, it is important to explore ways to extend the instrumental approach to more questions than it has been used to test thus far.

Researchers have recently begun conducting other sorts neurolinguistics research on scalar implicatures using such instrumental approaches to examine what happens the moment a scalar implicature could be realized (e.g., Barbet & Thierry, 2016, 2017). Barbet and Thierry (2016), for example, took advantage of the P300 component, which is known to be increased when a person detects a rare stimulus that they have been waiting for (a "target"). In one condition, participants were instructed to press a button every time they see a word in which the number of uppercase letters matches the meaning of the word (e.g., they should press a button for *ALL* and *tWO*, but not for *aLL* or *tWo*), and were instructed to treat "some" as meaning "some but not all". Thus, when they saw *SOME*, they should not press a button, since the amount of uppercase letters (all of them) does not match

the intended meaning of the word (some, but not all). Nonetheless, participants showed an increased P300 component in this condition, possibly suggesting that the brain initially interpreted these words as targets (i.e., initially interpreted "some" as meaning "at least one") despite instructions not to. Crucially, just like the example of van Turenhout and colleagues (1998) above, this study was not intended to reveal anything new about the P300 itself or to claim that the P300 is a locus of pragmatic processing; it simply used the P300 as an instrument to test specific theories of pragmatic processing.

Such approaches offer a fruitful way to link pragmatics and electrophysiology. Unlike the studies described above, the interpretation of the findings from these studies and their consequences for pragmatic theory is more straightforward, as they focus on components whose properties are well understood. When a well-understood component like a lateralized readiness potential, N400, or P300 is elicited, we generally know why it was elicited: because participants saw a word they were not expecting, saw a word that they need to press a button in response to, etc. It is then up to pragmatic theory and psycholinguistic models to explain why this happened. On the other hand, the interpretation of results from studies that are looking to observe neural correlates of pragmatic processing directly is generally problematic: if a study observes some brain component whose functional significance is unknown, then it is difficult to say what this means for theories of pragmatics, given that we don't know what that brain component means in of itself. Even for studies that observe known brain components, such as the N400 experiments reviewed above, they only are informative for theories of pragmatics if those components are interpreted instrumentally—as independent tools that give us insight into how much the pragmatic context caused a word to be expected, whether a scalar implicature was realized upstream, etc—rather than as themselves correlates of pragmatic processing.

While the instrumental approach is powerful, it of course has limitations. One is that many of these studies rely on unnatural tasks—e.g., the examples reviewed above used complex metalinguistic tasks that are quite different from naturalistic comprehension. Another limitation is that this approach requires both good instruments and good theories. Just as we would be unable to diagnose a fever well if we used a thermometer whose properties were not understood, a thermometer for which we do not know what will cause its reading to go up and down, we likewise cannot diagnose pragmatic theories if we rely on brain components that we don't know how to modulate. Likewise, a good instrument—a brain component with clear and predictable properties—is not useful if we cannot set up clear predictions for it: if the pragmatic theories and psychological models of pragmatic processing do not make clear and falsifiable predictions about what should happen under various circumstances, then instrumental approaches cannot unambiguously support or falsify them. Both of these limitations are potentially serious, given the state of pragmatics and neurolinguistics today: current models of pragmatic

processing are still mostly vague and each can generate or accommodate a wide variety of hypotheses, given a few degrees of freedom (Chemla & Singh, 2014a,b), and the exact functional interpretation of many event-related potential components is still under debate (for the N400, for example, see Lau et al., 2008). In other words, an experiment cannot adjudicate between two theories if both theories can claim to be able to accommodate the results of the experiment. Thus, instrumental approaches to neurolinguistics research on scalar implicatures are not applicable in all situations—it is certainly not the case that all pragmatics experiments in the future must use lateralized readiness potential or P300 designs! These instrumental approaches must be supplemented with other approaches, such as exploratory research in more natural settings, and psycholinguistic and linguistic research to help clarify the underlying models and the predictions they make. Nevertheless, instrumental approaches are an invaluable part of the neurolinguistics toolbox, and the study of scalar implicatures would benefit if they were more widely used.

Finally, another important consideration is what gains in knowledge about scalar implicature are uniquely attributable to the use of electrophysiological methods. As discussed above, if one accepts that the (or a) major goal of the neuro-pragmatics research enterprise is to understand the psychological mechanisms understanding pragmatics rather than to understand the brain itself, then instrumental research approaches are likely to be more productive than "modestly ontological" approaches. Much electrophysiological research on scalar implicatures is already using instrumental approaches, as acknowledged above; this can be exemplified by studies such as the study by Hunt and colleagues (2013) described above, and similar studies that examine how readers react to seeing a word that is unexpected because of a previously computed scalar implicature. In these studies, however, brain activity is being used as an instrument in much the same way behavioural data could be; while these studies may take advantage of the multidimensional data of electrophysiological data to search for different effects at different time windows or different parts of the scalp, they are not doing anything fundamentally different than what could be done by, e.g., measuring how quickly people read the unexpected words. Thus, it is difficult to point to them as illustrations of how electrophysiology can teach us about scalar implicature processing. Contrast these with studies that use the brain as an instrument in a way that behavioural methods cannot, such as the abovementioned examples from van Turrenout and colleagues (1998) and Barbet & Thierry (2016); both of these measure brain responses in the absence of observable behaviour. For researchers who are interested in avoiding the pitfalls of difficult-to-interpret results and futile searches for neural correlates of poorly understood psychological processes, but who still want to tap the power of electrophysiological methods and justify why such methods (as opposed to [usually] simpler behavioural methods) would be worthwhile for studying pragmatics and cognition, instrumental approaches that focus on paradigms unique to neuroscience would be fruitful.

Conclusion

Recent years have seen an explosion of research using neurolinguistics methods to study the processing of scalar implicatures. The relevance of much of this work to our understanding of how scalar implicatures occur and how they are processed, however, is still limited; in most cases, it is not possible to conclude that brain responses observed in scalar implicature experiment reflect scalar implicatures at all. While it is possible to learn about how scalar implicatures work without examining the neural patterns that emerge during the realization of the scalar implicature itself (e.g., by examining the downstream consequences of the scalar implicature), it is important for readers and researchers to be aware of the difference, and to not be under the illusions that using neuroscience methods automatically provides a window into how scalar implicatures emerge in the brain. The answer to the question raised in the title of this paper, then, is that electrophysiology can teach us very little about the cognitive processing of scalar implicatures unless future research can move beyond, or supplement, extant approaches which mainly either focus on an arguably misguided attempt to find "neural correlates" of pragmatic processes or which use electrophysiological methods to do the same things that behavioural methods can do. Researchers and students wishing to use electrophysiological methods and findings to shed light on how scalar implicatures work should interpret these results with caution, and are advised to consider adopting a more instrumental approach which would allow neurolinguistics experiments to more directly test theories of scalar implicature processing.

References

- Augurzky, P., Franke, M., & Ulrich, R. (2019). Gricean expectations in online sentence comprehension: an ERP study on the processing of scalar inferences. *Cognitive Science*, 43, e12776.
- Barbet, C., & Thierry, G. (2016). Some alternatives? Event-related potential investigation of literal and pragmatic interpretations of *some* presented in isolation. *Frontiers in Psychology*, 7, 1479. doi:10.3389/fpsyg.2016.01479
- Barbet, C., & Thierry, G. (2017). When 'some' triggers a scalar inference out of the blue. An electrophysiological study of a Stroop-like conflict elicited by single words. Poster presented at *7th biennial Experimental Pragmatics Conference*, Cologne, Germany.

- 724 Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the
725 comprehension of pragmatic inferences. *Journal of Experimental Psychology:*
726 *Learning, Memory, and Cognition*, 38, 1450. doi:10.1037/a0027850
- 727 Bezuidenhout, A. & Cutting, J.C. (2002). Literal meaning, minimal propositions, and
728 pragmatic processing. *Journal of Pragmatics*, 34, 433-456. doi:10.1016/S0378-
729 2166(01)00042-X
- 730 Bott, L., Bailey, T., & Grodner, D. (2012). Distinguishing speed from accuracy in
731 scalar implicatures. *Journal of Memory and Language*, 66, 123-142.
732 doi:10.1016/j.jml.2011.09.005
- 733 Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset
734 and time course of scalar inferences. *Journal of Memory and Language*, 51, 437-
735 457. doi:10.1016/j.jml.2004.05.006
- 736 Breheny, R., Ferguson, H., & Katsos, N. (2012). Investigating the time-course of
737 accessing conversational implicatures during incremental sentence interpretation.
738 *Language and Cognitive Processes*, 28, 443-467.
739 doi:10.1080/01690965.2011.649040
- 740 Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step:
741 Toward a model of on-line access to conversational implicatures. *Cognition*, 126,
742 423-440. doi:10.1016/j.cognition.2012.11.012
- 743 Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures
744 generated by default? An on-line investigation into the role of context in generating
745 pragmatic inferences. *Cognition*, 100, 434-463.
746 doi:10.1016/j.cognition.2005.07.003
- 747 Chemla, E., Cummins, C., & Singh, R. (2017). Training and timing local scalar
748 enrichments under global pragmatic pressures. *Journal of Semantics*, 34, 107-126.
749 Doi:10.1093/jos/ffw006
- 750 Chemla, E., & Singh, R. (2014a). Remarks on the experimental turn in the study of
751 scalar implicature, Part 1. *Language and Linguistics Compass*, 8, 373-386.
752 doi:10.1111/lnc3.12081
- 753 Chemla, E., & Singh, R. (2014b). Remarks on the experimental turn in the study of
754 scalar implicature, Part 2. *Language and Linguistics Compass*, 8, 387-399.
755 doi:10.1111/lnc3.12080
- 756 Chemla, E., & Spector, B. (2011). Experimental evidence for embedded scalar
757 implicatures. *Journal of Semantics*, 28, 359-400. doi:10.1093/jos/ffq023
- 758 Chevallier, C., Bonnefond, M., Van der Henst, J., & Noveck, I. (2010). Using ERPs
759 to capture inferential processes guided by prosodic cues. *Italian Journal of*
760 *Linguistics*, 22, 125-152.

- 761 Chevallier, C., Noveck, I., Nazir, T., Bott, L., Lanzetti, V., & Sperber, D. (2008).
762 Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology*, 61,
763 1741–1760. doi:10.1080/17470210701712960
- 764 Chierchia, G., Fox, D., & Spector, E. (2012). Scalar implicature as a grammatical
765 phenomenon. In: *Semantics: An International Handbook of Natural Language*
766 *Meaning*. Vol. 3. Edited by Maienborn, von Stechow & Portner. Mouton de
767 Gruyter. pp. 2297–2331.
- 768 De Neys, W., & Schaeken, W. (2007). When people are more logical under
769 cognitive load. *Experimental Psychology*, 54, 128–133. doi:10.1027/1618-
770 3169.54.2.128
- 771 Degen, J. (2015). Investigating the distribution of some (but not all) implicatures
772 using corpora and web-based methods. *Semantics and Pragmatics*, 8, 1–55.
773 doi:10.3765/sp.8.11
- 774 Degen, J. & Tanenhaus, M. (2015). Processing scalar implicature: a Constraint-
775 Based approach. *Cognitive Science*, 39, 667–710. doi:10.1111/cogs.12171
- 776 Dieussaert, K., Verkerk, S., Gillard, E., & Schaeken, W. (2011). Some effort for
777 some: Further evidence that scalar implicatures are effortful. *Quarterly Journal of*
778 *Experimental Psychology*, 64, 2352–2367. doi:10.1080/17470218.2011.588799
- 779 Doran, R., Baker, R., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-
780 unified nature of scalar implicature: an empirical investigation. *International*
781 *Review of Pragmatics*, 1, 211–248. doi:10.1163/187730909x12538045489854
- 782 Dwivedi, V., Phillips, N., Einagel, S., & Baum, Sh. (2010). The neural underpinnings
783 of semantic ambiguity and anaphora. *Brain Research*, 1311, 93–109.
- 784 Feeney, A., Scafton, S., Duckworth, A., & Handley, S. (2004). The story of some:
785 Everyday pragmatic inferences by children and adults. *Canadian Journal of*
786 *Experimental Psychology*, 54, 128–133. doi:10.1037/h0085792
- 787 Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on
788 object identification. *Cognitive Brain Research*, 16, 123–144. doi: 10.1016/S0926-
789 6410(02)00244-6
- 790 Geurts, B. (2010). *Quantity Implicatures*. Cambridge: Cambridge University Press.
- 791 Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures?!? *Semantics and*
792 *Pragmatics*, 2, 4–1. doi:10.3765/sp.2.4
- 793 Goodman, N., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling
794 language understanding as social cognition. *Topics in Cognitive Science*, 5, 173–
795 184. doi:10.1111/tops.12007

- 796 Goto, S., Ando, Y., Huang, C., Yee, A., & Lewis, R. (2010). Cultural differences in
797 the visual processing of meaning: detecting incongruities between background and
798 foreground objects using the N400. *SCAN*, 5, 242-253. doi: 10.1093/scan/nsp
- 799 Grice, H. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University
800 Press.
- 801 Grodner, D., Klein, N., Carbary, K., & Tanenhaus, M. (2010). "Some," and possibly
802 all, scalar inferences are not delayed: Evidence for immediate pragmatic
803 enrichment. *Cognition*, 116, 42. doi:10.1016/j.cognition.2010.03.014
- 804 Hartshorne, J., Liem Azar, S., Snedeker, J., & Kim, A. (2014). The neural
805 computation of scalar implicature. *Language, Cognition and Neuroscience*, 30, 620-
806 634. doi:10.1080/23273798.2014.981195
- 807 Hartshorne, J., & Snedeker, J. (manuscript). *The speed of inference: Evidence*
808 *against rapid use of context in calculation of scalar implicatures*.
- 809 Horn, L. (1972). *On the semantic properties of logical operators in English* (Ph.D.
810 dissertation). University of California, Los Angeles.
- 811 Huang, Y., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight
812 into the semantics-pragmatics interface. *Cognitive Psychology*, 58, 376-415.
813 doi:10.1016/j.cogpsych.2008.09.001
- 814 Hunt, L., Politzer-Ahles, S., Gibson, L., Minai, U., & Fiorentino, R. (2013). Pragmatic
815 inferences modulate N400 during sentence comprehension: Evidence from picture-
816 sentence verification. *Neuroscience Letters*, 534, 246-251.
817 doi:10.1016/j.neulet.2012.11.044
- 818 Kaan, E. (2007). Event-related potentials and language processing: a brief
819 overview. *Language and Linguistics Compass*, 1, 571-591. Doi:10.1111/j.1749-
820 818X.2007.00037.x
- 821 Katsos, N., & Cummins, C. (2010). Pragmatics: From theory to experiment and
822 back again. *Language and Linguistics Compass*, 4, 282-295. doi:10.1111/j.1749-
823 818X.2010.00203.x
- 824 Kowalewski, J., & Murphy, C. (2012). Olfactory ERPs in an odor/visual congruency
825 task differentiate ApoE ε4 carriers from non-carriers. *Brain Research*, 1442, 55-65.
826 doi: 10.1016/j.brainres.2011.12.030
- 827 Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:
828 (de)constructing the N400. *Nature Reviews Neuroscience*, 8, 920-933.
- 829 Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized*
830 *conversational implicature*. Cambridge: MIT press.
- 831 Marty, P., & Chemla, E. (2013). Scalar implicatures: Working memory and a
832 comparison with *only*. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00403

- 833 Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items
834 under memory load. *Lingua*, 133, 152–163. doi:10.1016/j.lingua.2013.03.006
- 835 Miller, D., & Rothman, J. (in press). You win some, you lose some: comprehension
836 and event-related potential evidence for L1 attrition. *Bilingualism: Language and*
837 *Cognition*.
- 838 Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of
839 pragmatic processing: An ERP investigation of informativeness and pragmatic
840 abilities. *Journal of Memory and Language*, 63, 324–346.
841 doi:10.1016/j.jml.2010.06.005
- 842 Noveck, I., & Posada, A. (2003). Characterizing the time course of an implicature:
843 An evoked potentials study. *Brain and Language*, 85, 203–210.
844 doi:10.1016/S0093-934X(03)00053-1
- 845 Noveck, I., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the
846 study of language. *Trends in Cognitive Sciences*, 12, 425–431.
847 doi:10.1016/j.tics.2008.07.009
- 848 Noveck, I., & Sperber, D. (2007). The why and how of experimental pragmatics:
849 The case of 'scalar inferences' In: Burton-Roberts N, editor. *Advances in*
850 *Pragmatics*. Basingstoke: Palgrave. pp. 184–212.
- 851 Panizza, D., & Onea, E. (2014). Some implicatures take their time: An ERP study on
852 scalar implicatures with 'sentence-picture vs. picture-sentence' verification task.
853 Talk presented at 27th CUNY Conference on Human Sentence Processing,
854 Columbia, SC, USA.
- 855 Panizza, D., Onea, E., & Mani, N. (2017). Early ERP evidence for children's and
856 adult's sensitivity to scalar implicatures. Talk presented at *Revising formal semantic*
857 *and pragmatic theories from a neurocognitive perspective*, Bochum, Germany.
- 858 Poeppel, D., & Embick, D. (2005) Defining the relation between linguistics and
859 neuroscience. In A. Cutler [Ed.], *Twenty-first century psycholinguistics: Four*
860 *cornerstones*. Lawrence Erlbaum.
- 861 Poldrack R. (2006). Can cognitive processes be inferred from neuroimaging data?,
862 *Trends in Cognitive Sciences*, 10, 59-63. doi: 10.1016/j.tics.2005.12.004
- 863 Politzer-Ahles, S. (2013). *Psycholinguistic and neurolinguistic investigations of*
864 *scalar implicature*. Doctoral dissertation, University of Kansas.
- 865 Politzer-Ahles, S. (2015). "Maybe" not all scalar implicatures are created equal. *LSA*
866 *Extended Abstracts*.
- 867 Politzer-Ahles, S. (manuscript). *MEG experiment on scalar implicature using simple*
868 *composition paradigm*. Unpublished manuscript. <https://osf.io/c9taq/>

- 869 Politzer-Ahles, S., & Fiorentino, R. (2013). The realization of scalar inferences:
870 Context sensitivity without processing cost. *PloS ONE*, 8(5), e63943.
871 doi:10.1371/journal.pone.0063943
- 872 Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural
873 correlates for pragmatic and semantic meaning processing: An event-related
874 potential investigation of scalar implicature processing using picture-sentence
875 verification. *Brain Research*, 1490, 134–152. doi:10.1016/j.brainres.2012.10.042
- 876 Politzer-Ahles, S., & Gwilliams, L. (2015). Involvement of prefrontal cortex in scalar
877 implicatures: evidence from magnetoencephalography. *Language, Cognition and*
878 *Neuroscience*, 30, 853-866. doi:10.1080/23273798.2015.1027235
- 879 Politzer-Ahles, S., & Husband, E. (in press). Eye movement evidence for context-
880 sensitive derivation of scalar inferences. *Collabra: Psychology*.
- 881 Rees, A., & Bott, L. (2018). The role of alternative salience in the derivation of
882 scalar implicatures. *Cognition*, 176, 1-14.
- 883 Rees, A., Bott, L., & Schumacher, P. (2019). Event-related potentials in pragmatic
884 priming. *Neuroscience Letters*, 712, 134435.
- 885 Sauerland, U. (2012). The Computation of scalar implicatures: Pragmatic, lexical or
886 grammatical? *Language and Linguistics Compass*, 6, 36–49. doi:10.1002/lnc3.321
- 887 Sauerland, U., & Schumacher, P. (2016). Pragmatics: theory and experiment
888 growing together. *Linguistische Berichte*, 245, 3-24.
- 889 Shetreet, E., Chierchia, G., & Gaab, N. (2014a). Linguistic inability or poor
890 performance: Dissociating scalar implicature generation and mismatch in the
891 developing brain. *Developmental Psychology*, 50, 2264-75. doi:10.1037/a0037368
- 892 Shetreet, E., Chierchia, G., & Gaab, N. (2014b). When three is not some: on the
893 pragmatics of numerals. *Journal of Cognitive Neuroscience*, 26, 854-63.
894 doi:10.1162/jocn_a_00514
- 895 Shetreet, E., Chierchia, G., & Gaab N. (2014c). When some is not every:
896 dissociating scalar implicature generation and mismatch. *Human Brain Mapping*, 35,
897 1503-14. doi:10.1002/hbm.22269
- 898 Sikos, L., Tomlinson, S., Traut, H., & Grodner, D. (2013, March). *Incremental*
899 *computation of scalar implicatures: An ERP study*. Poster presented at 26th CUNY
900 Conference on Human Sentence Processing, Columbus, OH, USA.
- 901 Spychalska, M., Kontinen, J., Noveck, I., Reimer, L., & Werning, M. (2019). When
902 numbers are not exact: ambiguity prediction in the processing of sentences with
903 bare numerals. *Journal of Experimental Psychology: Learning, Memory, and*
904 *Cognition*, 45, 1177-1204.

- 905 Spychalska, M., Kontinen, J., Werning, M. (2016) Investigating scalar implicatures
906 in a truth-value judgment task: Evidence from event-related brain potentials.
907 *Language, Cognition and Neuroscience*, 31, 817-840.
908 doi:10.1080/23273798.2016.1161806
- 909 Spychalska, M., Reimer, L., P. Schumacher, & Werning, M. (2017). Scalar
910 implicatures in the context of full and partial information: evidence from ERPs.
911 Poster presented at *7th biennial Experimental Pragmatics Conference*, Cologne,
912 Germany.
- 913 Tomlinson, J., Bailey, T., & Bott, L. (2013). Possibly all of that and then some:
914 scalar implicatures are understood in two steps. *Journal of Memory and Language*,
915 69, 18-35. doi:10.1016/j.jml.2013.02.003
- 916 van Berkum, J. (2009). The neuropragmatics of 'simple' utterance comprehension:
917 an ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and Pragmatics:*
918 *From Experiment to Theory* (pp. 276-316). Basingstoke: Palgrave Macmillan.
- 919 van Berkum, J. (2010). The brain is a prediction machine that cares about good and
920 bad. *Rivista di Linguistica*, 22, 181-208.
- 921 van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2016). Scalar
922 diversity. *Journal of Semantics*, 33, 107-135. doi:10.1093/jos/ffu017
- 923 van Turenout, M., Hagoort, P., & Brown, C. (1998). Brain activity during speaking:
924 from syntax to phonology in 40 milliseconds. *Science*, 280, 572-574.
- 925 Virtue, S., & Sundermeier, B. (2016). What can neuroimaging research tell us
926 about inference generation during text comprehension? *Language & Linguistics*
927 *Compass*, 10, 257-271.
- 928 Zhao, M., Liu, T., Chen, G., & Chen, F. (2015). Are scalar implicatures automatically
929 processed and different for each individual? A mismatch negativity (MMN) study.
930 *Brain Research*, 1599, 137-149. doi:10.1016/j.brainres.2014.11.049

Study	Language	Scale	Manipulation	Position	Task	Effect family
Noveck & Posada (2003)	French	<some, all>	Violation of SI-based prediction	downstream	Metalinguistic judgment	reduced N400
Nieuwland et al. (2010)	English	<some, all>	Violation of SI-based prediction	scalar and downstream	Passive reading	N400 downstream, positivity on scalar
Chevallier et al. (2010)	French	<or, and>	Contextual support for inference	scalar	Metalinguistic judgment	P600
Politzer-Ahles et al. (2013)	Mandarin	<some, all>	Infelicitous scalar expression	scalar	Mixed	Sustained posterior/broad negativity
Politzer-Ahles (2013)	Mandarin	<some, all>	Infelicitous scalar expression	scalar	Metalinguistic judgment	Sustained broad negativity
Hunt et al. (2013)	English	<some, all>	Violation of SI-based prediction	downstream	Metalinguistic judgment	N400
Sikos et al. (2013)	English	<some, all>, <many, all>	Violation of SI-based prediction	scalar and downstream	Metalinguistic judgment	Anterior negativity on scalar, N400 downstream
Panizza et al. (2014)	German	<some, all>	Infelicitous scalar expression	scalar	Metalinguistic judgment	Left anterior negativity
Hartshorne et al. (2014)	English	<some, all>	Contextual support for inference	scalar and downstream	Comprehension questions	No difference at scalar; sustained anterior negativity downstream

Politzer-Ahles & Gwilliams (2015)*	English	<some, all>	Contextual support for inference	scalar	Comprehension questions	Sustained frontal lobe activation
Politzer-Ahles (2015)	English	<maybe, definitely>	Infelicitous scalar expression	scalar	Metalinguistic judgment	N400
Politzer-Ahles (ms.)*	English and Arabic	<some, all>	Contextual support for inference	scalar	Picture-word matching	No robust pattern
Zhao et al. (2015)	Mandarin	<some, all>	Violation of SI-based prediction	downstream	Passive listening	MMN-like
Spychalska et al. (2016)	German	<some, all>	Violation of SI-based prediction	downstream	Metalinguistic judgment	N400
Barbet & Thierry (2016)	English	<some, all>	Oddball	scalar	Target detection	P3b
Panizza et al. (2017)	German (child & adult)	<some, all>	Infelicitous scalar expression	scalar	Metalinguistic judgment	Sustained negativity
Spychalska et al. (2017)	German	<some, all>	Infelicitous scalar expression	downstream	Metalinguistic judgment	Negativity
Barbet & Thierry (2017)	English	<some, all>	Stroop	scalar	Stroop judgment	N450
Augurzký et al. (2019)	German	<some, all>	Infelicitous scalar expression, violation of SI-based prediction	scalar and downstream	Metalinguistic judgment	positivity at quantifier, N400 downstream

Rees et al. (2019)	English	<some, all>; <or, and>	Pragmatic priming	downstream picture verification	Metalinguistic judgment	P300/P600
Spychalska et al. (2019)	German	bare numerals	Violation of SI-based prediction	downstream	Metalinguistic judgment	N400 / sustained negativity
Miller & Rothman (in press)	Spanish	<some, all>	Infelicitous scalar expression	scalar	Metalinguistic judgment	N400

*Table 1. Summary of extant scalar implicature studies using electrophysiology. Studies using magnetoencephalography are indicated with a *; studies used electroencephalography unless otherwise indicated.*

Figure captions

Figure 1. Sample stimuli from Hunt et al. (2013).

Figure 2. Sample stimuli from Politzer-Ahles et al. (2013).