CHAPTER 3

IS SELF-PACED LISTENING SENSITIVE TO DOWNSTREAM CONSEQUENCES OF FOCUS?

STEPHEN POLITZER-AHLES
JULIE S. CHEN
I-HSUAN CHEN

1. Introduction

The word *even*, and equivalents in other languages, suggests that the thing being described is very unlikely, and thus it invokes a scale of more- and less-likely things. For example, a person uttering (1a) might be implicating that a dog is the most likely thing someone might see at a park, and thus if they didn't see a dog then it follows that they also didn't see an alligator, a unicorn, etc. (1b).

(1a) At the park I didn't even see a dog.

(1b) At the park I didn't even see a dog, much less an alligator (or any other animal that's less likely to be seen at a park than a dog is)

This scale-invoking behavior is traditionally explained as being not part of the literal meaning of *even*, but rather coming from an implicature. In other words, the scale suggested by *even* does not affect the truth or falsity of the sentence, but comes from the fact that *even* conventionally suggests the speaker's belief that the proposition describes something very unlikely.[1]

[1] This interpretation is usually viewed as a conventional implicature, as described here (see, e.g., Francescotti 1995). See, however, Boguslavsky (2001) for an argument that it is instead a conversational implicature. This distinction, while important for a theory of pragmatics, will not have an impact on the details or predictions of the present study.

Hence a sentence like (2) is not false, even though the scale it suggests is false (at least in most ordinary contexts, where Martians are not the most likely thing for someone to see at the park); it is, however, weird-sounding and infelicitous, unless a particular context is given.

(2) At the park I didn't even see a Martian.

Figuring out precisely what scale an utterance with *even* is meant to evoke, however, may pose a challenge for the language comprehension system (even though people normally accomplish this rapidly and effortlessly in natural communication, just like most implicatures). For a sentence like (1a), the implicature about what is the most likely proposition could target almost any part of the sentence, as shown in (3) below:

(3a) At the park I didn't even see a **dog**, much less an alligator.

(3b) At the park I didn't even **see** a dog, much less roll around in the grass with one.

(3c) At the park **I** didn't even see a dog – so my friend who was looking at his phone all the time surely didn't, either.

In (3a), the implicature is that dogs are one of the most likely things to be seen in the park, and thus the speaker didn't see other, less likely things. In (3b), on the other hand, the implicature is that seeing a dog is one of the most likely things one would *do* with a dog at the park, and thus the speaker didn't do other, less likely activities. And in (3c), the implicature is that the speaker is one of the most likely people to see a dog (perhaps in this context the speaker has been specifically trying to spot dogs), and thus other people who are less likely to see a dog (perhaps because they are paying attention to something else) also didn't see any dogs.

Incidentally, other scope operators, such as *only*, also show this behavior; for example, (4) can be interpreted as (4a) or as (4b):

(4) I only pet the dog.

(4a) I only pet the **dog** (not the cat, the pony, etc.).

(4b) I only **pet** the dog (but I didn't feed it, wash it, take it for a walk, etc.)

1.1 Focus and the interpretation of even

Given this ambiguity, how do interlocutors accurately recover which implicature is intended by the utterer? Context, of course, plays a large role; for example, if someone utters (1a) in response to a question about what kinds of animals he saw at the park, then the question under discussion in the discourse already constrains which implicature is relevant. Over and above the effect of context, a wide variety of linguistic devices are used across languages to disambiguate focus (Gussenhoven 2008). One of these is prosody. If the constituent that represents the most likely member of the scale being invoked (the bolded words in (3)) is pronounced with phonological prominence (stress) leading to narrow focus on that constituent (Ladd 1980), then it seems easier to recover the intended implicature about that constituent. In other words, it seems easier to recover the implicature "A dog is the most likely thing to see at the park" from utterance (3a), with stress on "dog", than from (3b) or (3c), with stress on other words.

Chen and colleagues (2018) describe a series of experiments in Mandarin which might support this claim. Participants listened to the sentence fragments like (5a) and (5b), in a context describing a cat café that is closed for the day:

(5) Auditory sentence fragments with *even*

(5a) **Noun stress:** 連 一隻 **貓咪** 都沒有。
even one cat there isn't
"There isn't even one **cat**."

(5b) **Num+CL stress:** 連 **一隻** 貓咪 都沒有。
even one cat there isn't
"There isn't even **one** cat."

In a given trial, participants listened to both of these and were shown one of the two sentence frames from (6) (the gloss SFP stands for "sentence-final particle"):

(6) Written sentence frames

(6a) **Type contrast:**
___, 更別說 有 顧客 了。
not to mention there are customers SFP
"___, not to mention customers."

(6b) **Quantity contrast:**

___,	更別說	有	一群貓咪	了。
	not to mention	there are	a group of cats	SFP

"___, not to mention a group of cats."

After seeing one sentence frame and hearing both versions of the audio, participants were asked to choose which audio better fits the sentence frame. When the sentence frame invoked a contrast about which type of things one might see in a cat café (6a), participants were more likely to choose the audio sentence fragment with stress on the noun *cat* (5a); in other words, people preferred "*There isn't even one **cat**, not to mention any customers*" over "*There isn't even **one** cat, not to mention any customers*". Likewise, when the sentence frame invoked a contrast about how many cats might be in a cat café (6b), participants were more likely to choose the audio sentence fragment with stress on the numeral-classifier phrase *one* (5b). In other words, the location of the prosodic stress on one or another constituent in the clause with *even* seemed to influence listeners' interpretation of what implicature, invoking what scale, should be recovered from the utterance.

Several limitations of this study motivate the need for further investigation, though. First of all, the evidence itself in this study is rather weak: the first two experiments in the study failed to find effects of prosody on the interpretation of sentences with *even*, and only the third experiment, described above, managed to observe such effects. Therefore, it is worth seeking converging evidence from other methods to see whether there really are robust effects of prosody on the interpretation of these sorts of sentences, i.e., to rule out the possibility that the effect reported above is a lone fluke. Furthermore, the study used a fairly unnatural and metalinguistic task, and thus was not able to test whether listeners actively use prosody in real-time during natural conversation to constrain the kinds of inferences that could be recovered from utterances with *even*. The present study attempts to address both of these limitations by testing listeners' interpretation of clauses like (5a, 5b), this time using the psycholinguistic technique of *self-paced listening*.

1.2 Self-paced listening

In self-paced listening (Ferreira *et al.* 1996; see Papadopoulou *et al.* 2013, for a review), participants hear a recording of a sentence, divided up into smaller segments (words or phrases). They advance through the sentence by pressing a button whenever they are ready to proceed to the next segment. By measuring how long they take to press the button when listening to any given segment, it is possible to infer how long it took them to process that

segment (i.e., to comprehend its meaning, integrate it into the surrounding sentence and/or discourse context, predict what's coming next, and do whatever else is involved in sentence comprehension). It is a variant of the self-paced reading paradigm (Just, Carpenter and Wooley 1982), in which participants read a sentence presented on the screen one segment at a time by pressing a button to proceed to each successive segment. Self-paced reading is much more widely used than self-paced listening, given that presenting stimuli by text is much easier and more efficient (not to mention that it allows more precise control of timing) than preparing and editing audio recordings. Nevertheless, self-paced listening is a useful technique for investigating language comprehension in, for example, speakers of languages that are not commonly written (see, e.g., Wagers, Borja and Chung 2015) and other non-reading populations. It is also useful when spoken language is necessarily the target of investigation, as with research on prosody. This makes self-paced reading an appropriate method to test how stress is used in real time to recover implicatures, as discussed above.

To the best of our knowledge, the vast majority of self-paced listening research has focused on parsing syntax, particularly on phenomena such as the garden path effect and the resolution of attachment ambiguities. A study by Papadopoulou and colleagues (2015) used self-paced listening to examine reference resolution, which is a topic closer to pragmatics (although it is the type of pragmatics sometimes called "semantic pragmatics" or "the pragmatics of what is said" [Recanati 1989], as opposed to the pragmatics of implicature). We are not aware of any studies using self-paced listening to examine the processing of pragmatic implicatures, particularly ones that examine how prosody impacts the recovery of pragmatic implicatures. In fact the influence of prosody on downstream self-paced listening time is unclear; for example, in the original self-paced listening study by Ferreira and colleagues (1996), prosody failed to prevent participants from experiencing a garden-path effect. This stands in contrast to other techniques such as the visual world eye-tracking paradigm, where studies have repeatedly and robustly found that prosody is used to avoid ambiguity (e.g. Snedeker and Trueswell 2003; for review see Huettig, Rommers and Meyer 2011). Thus, while self-paced listening may be useful for addressing this study's own research question, this study may also be useful to provide further information on whether self-paced listening is sensitive to the online use of prosody to recover pragmatic implicatures (assuming prosody is indeed used online to do this).

1.3 The present study

The aim of the present study is to use self-paced listening to test whether prosody moderates the implicatures that are recovered from an utterance with *even*, as evidenced by how these implicatures do or do not facilitate listening times to a downstream word.

Specifically, we had participants listen sentences which began with clauses like those in (7), with stress either on a noun (7a) or a numeral-classifier phrase (7b). Both versions of the sentence ended with clauses like (8).

(7) Initial sentence fragments for the present study

(7a) **Noun stress:** 学校里 **连** 一个 **学生** 都没来
at school even one student didn't come
"At school there wasn't even one **student**…"

(7b) **Num+CL stress:** 学校里 **连** **一个** 学生 都没来
at school even one student didn't come
"At school there wasn't even **one** student…"

(8) Critical clause: 更不用**说** 有 老**师** 在走**动**了
not to mention there are professors walking around
"…not to mention professors."

If prosody constrains the implicatures recovered from (7) on-line, then we expect that the listening time for "professors" later in the sentence (i.e., in (8)) will be faster if the sentence began with stress on the noun (7a) than if the sentence began with stress on the numeral-classifier phrase (7b). Specifically, in (7a), stress on the noun should help the listener recover the implicature that students are the most common type of people hanging around at school, and from there they can infer that other less common types of people are also not there; then, when they encounter "not to mention…", they should predict that the upcoming word will be something from that set of less-common types of people, such as professors. On the other hand, in (7b), stress on the numeral-classifier phrase should lead readers to recover a different implicature, i.e. that the most likely number of students that would be at school is at least one, and so if that number of students was not at school then a larger number was also not at school. In this case the implicature would not particularly facilitate the comprehension of "professors" later.

The design of this experiment presupposes that recognition of an implicature can facilitate downstream processing of a later word or expression that that implicature helps the listener predict. We believe this is a fair assumption, though, as it has been widely shown in other manipulations. For instance, conversational implicatures have been shown to facilitate comprehension of a later word in self-paced reading (most famously by Breheny, Katos and Williams 2006; see Politzer-Ahles and Husband 2018, for a review of other studies using their paradigm), eye-tracking while reading (Politzer-Ahles and Husband 2018), and event-related potentials (Hunt *et al.* 2013; Spychalska, Kontinen and Werning 2016). Most relevant to the present study, a recent eye-tracking while reading experiment (Ivanova and Bello Viruega 2019) also suggests that implicatures associated with *even* facilitates downstream reading times. In this study, participants read sentences (9a-b), and the final word *punk* was read numerically more quickly (in first-pass time, second-pass time, and total time) if it was preceded by *even* (9b) than if it was preceded by *and* (9a). While the study is preliminary (the part of the study about *even* only includes one item, which was repeated twice [once per condition] within each participant; and the authors do not report statistical comparisons between the same word preceded by different connectives), the results are consistent with the notion that *even* triggered an implicature which led readers to expect that the next word would be a kind of music that people who like folk/soul/jazz are normally unlikely to enjoy, and that readers used that expectation to facilitate their processing of the next word *punk* (e.g. by pre-constraining the set of which types of music they predict), compared to the sentence with *and*, which does not offer this sort of indication about what the next type of music in the list will be.

(9a) Mike and Lucy love folk, soul, jazz and punk.

(9b) Mike and Lucy love folk, soul, jazz, even punk.

Therefore, we believe the existing literature convincingly shows that implicatures can modulate the speed at which later words are comprehended. The present study takes advantage of this fact to use the comprehension speed of a downstream word as a probe for testing whether or not prosody modulated the recovery of that implicature earlier on.

2. Methods

2.1 Preliminary experiment

We ran a preliminary version of the experiment described above, with 20 participants, but do not report the details of that experiment here. In the preliminary version, we set up the experiment script such that no matter when participants pressed the button to continue, the present audio segment played to the end before the program proceeded to the next audio segment. Because of this procedural flaw, participants generally just repeatedly mashed the 'continue' button as fast as they could – no matter how much or how quickly they mashed the button, the sentence would still play in full. Therefore, that version of the experiment did not meaningfully measure processing speed. We therefore re-ran the experiment with the same stimuli (but new participants) in a different procedure, where pressing a button before a segment was finished would end playback of that segment and proceed to the next segment; this forced participants to only press the button when they had comprehended the current segment and were ready to hear the next one. This second version of the experiment is what we report below.

2.2 Participants

We collected data from 44 native speakers of Mandarin (average age 25, range 19-36, 18 men and 26 women; see https://osf.io/8nvm7/ for detailed demographic information). Procedures for the experiment were approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University. All participants provided written informed consent and were paid for their participation.

2.3 Materials

The critical materials comprised 24 stimulus sets; one example stimulus set is shown in (10). Slashes indicate where the audio recording was divided into segments.

(10a) **Noun stress, type contrast**
学校里 / 连一个**学生** / 都没来 / 更不用说 / 有老师 / 在走动 / 了
at school / even one **student** / didn't come / not to mention / professors / walking around / SFP

(10b) **Num+CL stress, type contrast**

学校里 / 连**一个**学生 / 都没来 / 更不用**说** / 有老**师** / 在走**动** / 了
at school / even **one** student / didn't come / not to mention / professors / walking around / SFP

(10c) **Noun stress, quantity contrast**
学校里 / 连一个**学生** / 都没来 / 更不用**说** / 有一群学生 / 在 / 了
at school / even one **student** / didn't come / not to mention / a group of students / there / SFP

(10d) **Num+CL stress, quantity contrast**
学校里 / 连**一个**学生 / 都没来 / 更不用**说** / 有一群学生 / 在 / 了
at school / even **one** student / didn't come / not to mention / a group of students / there / SFP

(10a-b) are the critical conditions, which are identical to what was described above in (7-8). (10c-d) are included just to make sure participants cannot predict "professors" coming up in the sentence. We did not analyze the reading times for these sentences, as we felt they sound fairly unnatural. The critical words used for the type contrast conditions (10a, 10b) were the most frequently chosen words for these sentence frames in a written sentence completion norming test. Specifically, we presented 50 participants (none of whom went on to participate in this study) with written versions of the sentence frames up to and including 更不用**说** ("not to mention"), and asked each participant to complete each sentence; participants were allowed to write down more than one possible completion for each sentence. The type-contrast critical words used in the eventual 24 stimulus sets all had cloze probabilities above 50%. The critical words were not all unique, though; in order to ensure high cloze probability for each critical word while also having sufficient items to test, some critical words were repeated across more than one item.

We prepared the audio stimuli in a way to ensure that the critical segment (e.g. "*professors*") would be the same physical stimulus across the conditions being compared (10a and 10b). First, the sentences were spoken aloud by a native Mandarin speaker, who produced all four versions of each item. The recorded stimuli were then manually segmented using Praat (Boersma and Weenink 2018). The noun-stressed and Num+CL-stressed segments (i.e., "*even one **student***" and "*even **one** student*" in (10)) were cut from two separate recordings, as were the critical segments including the contrast set and the following segment (e.g., *professors* or *a group of students* and the following segment in (10)). The other segments were all cut from the same recording. The final stimuli were created by combining

these separate segments as illustrated in Table 1 (these segments still remained in their own sound files, but were played in the order shown below by the experiment control software; see section 2.4). As shown in the table, the critical segments ("professors" and the following segment) are the same physical tokens across the two conditions we are interested in comparing; the only physical difference between conditions occurs several segments earlier than the critical segment.

Table 1.

	At school	even one **student** / even **one** student	didn't come	not to mention	professors / a group of students	walking around / there	SFP
10a	学校里 10a	连一个**学生** 10a	都没来 10a	更不用说 10a	有老师 10a	在走动 10a	了 10a
10b	学校里 10a	连**一个**学生 10b	都没来 10a	更不用说 10a	有老师 10a	在走动 10a	了 10a
10c	学校里 10a	连一个**学生** 10a	都没来 10a	更不用说 10a	有一群学生 10c	在 10c	了 10a
10d	学校里 10a	连**一个**学生 10b	都没来 10a	更不用说 10a	有一群学生 10c	在 10c	了 10a

Illustration of how the recorded materials were combined into the final stimuli sentences, using the item shown in (10) as an example. Each row shows how one version of the item was constructed. The various columns of the row indicate which recording each segment was taken from. For example, (10b) was made up of segments cut from the recording of (10a), except for the second segment ("even one **student**" in (10a), but "even **one** student" in (10b)).

In addition to the critical stimuli described above, the experiment also included 35 filler stimuli to try to mask the purpose of the experiment and to prevent participants from always being able to predict the critical word. 12 of the fillers had the same structure as the critical stimuli, but the word in the critical position (i.e. the position that "*professors*" occupies in (10)) was completely unrelated to the context and should be completely unexpected, e.g., "*Today the factory was very empty, there was not even one worker, not to mention any monks*" (the word corresponding to the critical position, *monks*, is underlined). Each of these fillers was recorded in two versions, one with stress on the noun (as in (10a, 10c)) and one with stress on the numeral-classifier phrase (as in (10b, 10d)). The other 23 fillers were garden-path sentences taken from an unrelated experiment about relative clauses (Wu, Kaiser and Anderson 2011).

The full list of stimuli, as well as the recordings, are available at https://osf.io/8nvm7/.

2.4 Procedure

As the critical stimuli fell into four conditions, they were organized into six lists following a Latin square design (the 12 fillers that matched the structure of the critical items fell into two conditions, and were thus organized into two lists; thus, Latin square lists 1 and 3, for instance, had the same fillers. The other 24 fillers were the same across all lists.)

Experiment control and logging of responses and response times was handled by DMDX (Forster and Forster 2003); the DMDX scripts are available at https://osf.io/8nvm7/ (the scripts named "List{1-4}b.rtf"). Within each list, the 59 stimuli (24 critical + 35 filler) were grouped into five groups of ten trials each and one group of nine trials, such that each group included four critical stimuli, two of the fillers that matched the structure of the critical items, and 3-4 of the unrelated fillers. The order of the trials was pseudorandomized such that the order of the six groups was randomized and the order of the 9-10 trials within each group was randomized. Participants heard a segment over headphones, and pressed the space bar when they were ready to continue. When the participant pressed the space bar or when 30 seconds elapsed, the next segment was presented (stopping playback of the present segment, if it had not finished yet). After the sentence had finished playing, participants were shown a two-alternative multiple-choice question for comprehension on the screen, and had to choose one of the responses shown on the screen by pressing the left or right shift key. The location of the correct response varied across items. The experiment was preceded by a practice session of three trials.

3. Results

Data are available at https://osf.io/8nvm7/ (in the folder labelled "Exp2 data"; "Exp1 data" is the preliminary experiment described at the beginning of section 2). Accuracy on comprehension questions was high across the board (97% for trials with quantity contrast [10c-d] regardless of stress position, and 98% for trials with type contrast [10a-b] regardless of stress position).

Mean listening times per segment in the latter, critical, part of the sentence are shown in Figure 1. As discussed above, we expected that participants would spend less time listening to the critical noun (e.g. "professors") when the preceding clause had stress on the alternative noun

(e.g. “student”) rather than when the preceding clause had stress on the numeral-classifier phrase (“one”). This clearly was not the case; the listening for this segment times are almost identical (1247 ms for the noun-stress condition and 1249 ms in the numeral-classifier-stress condition) and these were not significantly different (b=1.6, t=0.06, p=.953).[2] Figure 2 shows the within-participant context effects (i.e., for each participant, the listening time to the critical word in the numeral-classifier-stress context minus the listening time to the same word in the noun-stress context; this difference was predicted to be positive), which helps us see that the failure to find a significant effect was not driven by a few outliers, but rather that across the whole dataset there was indeed no reliable trend one way or the other. We also do not see the predicted effect at the spillover segment (if anything, the listening time for this segment is numerically longer in the numeral-classifier focus condition, whereas we expected that condition would lead to shorter times). Thus, it is clear that, no matter how we look at it, the predicted effect is not present.

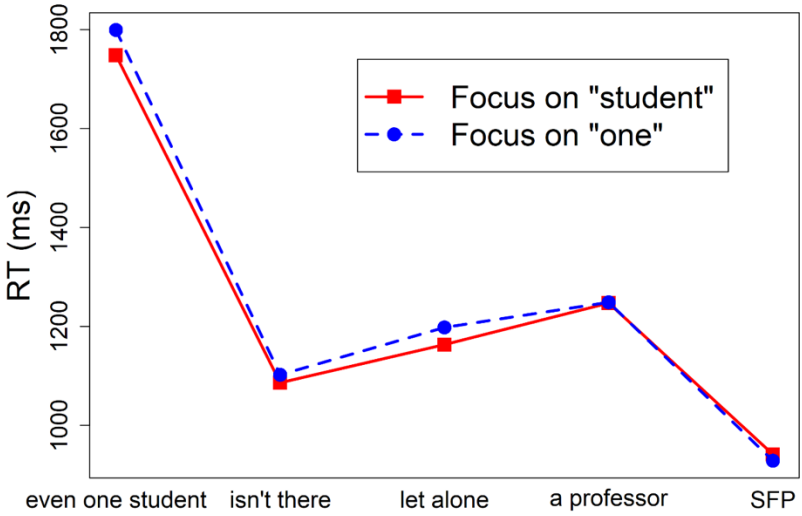
4. Discussion

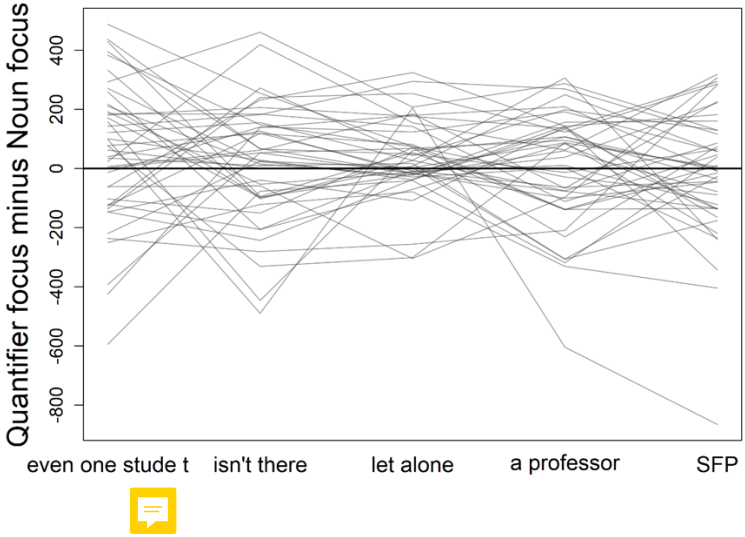
Contrary to our expectation, listening times to the critical word were not modulated by stress earlier in the sentence; in other words, we found no evidence that stress influenced the kind of implicature listeners recovered.

We can think of four possible explanations for this result. First of all, maybe stress and focus simply do not modulate how people infer scales related to *even* on-line. This seems unlikely, given that stress is widely acknowledged to have substantial impact on both end-state interpretation (e.g., Gussenhoven 2008) and on online processing as evidenced by techniques such as visual world eye-tracking (e.g. Snedeker and Trueswell 2003) and event-related potentials (e.g. Chevallier *et al.* 2010). Nevertheless, it is a possibility which must be acknowledged. There is some converging evidence; e.g., Lauter (2013) did not find modulation of scalar implicatures (as measured by modulation of reading times for a downstream word) by stress presented orthographically (through capital letters).

[2] The statistical analysis was done with linear mixed-effects models (Baayen, Davidson, and Bates 2008) implemented in the {lme4} package (Bates *et al.* 2015) of the R statistical computing environment (R Core Team 2016). Models with maximal random effects (Barr *et al.* 2013) were attempted but yielded singular fit, and simplifying them by removing random intercepts and intercept-slope correlation terms did not fix this, so the final models use only random intercepts. p-values are estimated using the {lmerTest} package (Kuznetsova, Brockhoff, and Christensen 2017) per the recommendation of Luke (2017).

Figure 1. Mean reading times per segment in the critical portion of the sentence. At the critical word, “professor”, reading times are roughly the same between the two conditions.





Secondly, maybe stress does influence what scales people infer from a conventional implicature, but these scales do not facilitate their downstream predictions or the ease with which they interpret later words. This also seems unlikely, based on the evidence reviewed in section 1.3.

Third, maybe self-paced listening is not sufficiently sensitive to the dynamics of sentence comprehension, at least those related to pragmatic processes. This is challenged by the fact that self-paced listening has been widely shown to elicit psycholinguistic effects that are comparable to those found in self-paced reading (see, however, the discussion in section 1.2 regarding the insensitivity of parsing to prosody in Ferreira and colleagues' (1996) self-paced listening experiment; see also Papadopoulou *et al.* (2013, 55-56) for a summary of contradictory findings in the literature). Attributing the present study's null effect to nebulous problems with "pragmatic processes" is also questionable, because the study was intentionally designed to not need to measure pragmatic processes directly, but to measure their downstream consequences, i.e., how hard a word is to comprehend later in the sentence. It is not clear why a word that's unexpected or hard to comprehend because of a pragmatic implicature or because of prosody would have different effects on self-paced listening than a word that's unexpected or hard to comprehend because of, say, a garden-path effect or an attachment ambiguity. If this is the case, though, that might suggest that reaction times in self-paced listening mainly reflect some other process (such as revision) that might more plausibly be claimed to be different in these cases. This is a question worth further study (although such study would also owe us an explanation of why this dissociation emerges in self-paced listening if it does not in self-paced reading). It could be valuable to examine this with measures that are potentially more sensitive and that can detect qualitatively different responses for different processes, such as event-related potentials (although the interpretation of event-related potential data comes with many of its own challenges; see Luck (2014), for general discussion, and Politzer-Ahles (2020), for discussion particular to pragmatics). Alternatively, a visual analogue of this manipulation could be tested by using capital letters or italic or boldfaced text, which would allow the study to be done in self-paced reading (the properties of which are better understood than self-paced listening) or in eye-tracking (which, like event-related potentials, brings the possibility of detecting qualitatively different responses for different cognitive processes, since eye-tracking allows many dependent measures).

Finally, it is of course possible that some other methodological flaw in our experiment prevented us from detecting the intended effect. For

instance, the repetition of some critical words across more than one sentence may have weakened the impact of the experimental manipulation on critical word predictability. Furthermore, since we chose critical words based on their cloze probability in the sentence contexts, we ended up using some critical words that do not have an obvious scalar relationship with the earlier part of the sentence; for instance, one item was "The shoe cabinet was just installed, there's not even one pair of slippers inside, not to mention leather shoes", and it's not obvious that slippers and leather shoes are in a scalar relationship in terms of which one is most likely to be found in a shoe cabinet. It is possible that items like these may have diluted the effectiveness of the experimental design. While we do not have a definitive answer to what might be the flaw (if any) that prevented the experiment from detecting the effect we had predicted, it must be acknowledged that the design of the present study is complicated and hinges on many assumptions (assumptions about what people infer from the clause with *even*, what contrast set that leads them to think of, and how that contrast set influences downstream processing) and each of these assumptions introduces a place where the experiment design may break down.

References

Baayen, Harald, Davidson, Doug, and Bates, Douglas. 2008. "Mixed-effects modeling with crossed random effects for subjects and items". *Journal of Memory and Language* 59: 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, Dale, Levy, Roger, Scheepers, Christoph, and Tily, Harry. 2013. "Random effects structure for confirmatory hypothesis testing: keep it maximal". *Journal of Memory and Language* 68: 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, Douglas, Maechler, Martin, Bolker, Ben, and Walker, Steve. 2015. "Fitting linear mixed-effects models using lme4". *Journal of Statistical Software* 67: 1-48. https://doi.org/10.18637/jss.v067.i01

Boersma, Paul, and Weenink, David. 2018. *Praat: doing phonetics by computer* [Computer program]. http://www.praat.org/

Boguslavsky, Igor. 2001. "On the scales and implicatures of *even*". In *Current Research in the Semantics/Pragmatics Interface: Volume 8* (*Pragmatics and The Flexibility of Word Meaning)*, edited by Eniko Nemeth and Károly Bibok, 29-50. Oxford: Elsevier.

Breheny, Richard, Katsos, Napoleon, and Williams, John. 2006. "Are generalised scalar implicatures generated by default? An on-line

investigation into the role of context in generating pragmatic inferences". *Cognition* 100: 434-463. https://doi.org/10.1016/j.cognition.2005.07.003

Chen, I-Hsuan, Huang, Chu-Ren, and Politzer-Ahles, Stephen. 2018. "Determining the types of contrasts: the influences of prosody on pragmatic inferences". *Frontiers in Psychology* 9: 2110. https://doi.org/10.3389/fpsyg.2018.02110

Chevallier, Coralie, Bonnefond, Mathilde, Van der Henst, Jean-Baptiste, and Noveck, Ira. 2010. "Using ERPs to capture inferential processes guided by prosodic cues". *Italian Journal of Linguistics* 22: 125-152. https://www.italian-journal-linguistics.com/app/uploads/2021/05/7_chevallier-et-alii.pdf

Ferreira, Fernanda, Henderson, John, Anes, Michael, Weeks, Philip, and McFarlane, David. 1996. "Effects of lexical frequency and syntactic complexity in spoken language comprehension: evidence from the auditory moving window technique". *Journal of Experimental Psychology: Learning, Memory & Cognition* 22: 324–335. https://doi.org/10.1037/0278-7393.22.2.324

Forster, Kenneth, and Forster, Jonathan. 2003. "DMDX: a Windows display program with millisecond accuracy". *Behavior Research Methods, Instruments, & Computers* 35: 116-124. https://doi.org/10.3758/BF03195503

Francescotti, Robert. 1995. "*Even*: the conventional implicature approach reconsidered". *Linguistics and Philosophy* 18: 153-173. https://doi.org/10.1007/BF00985215

Gussenhoven, Carlos. 2008. "Notions and subnotions in information structure". *Acta Linguistica Hungarica* 55, 381-395. http://dx.doi.org/10.1556/ALing.55.2008.3-4.11

Huettig, Falk, Rommers, Joost, and Meyer, Antje. 2011. "Using the visual world paradigm to study language processing: a review and critical evaluation". *Acta Psychologica* 137, 151-171. https://doi.org/10.1016/j.actpsy.2010.11.003

Hunt, Lamar, Politzer-Ahles, Stephen, Gibson, Linzi, Minai, Utako, and Fiorentino, Robert. 2013. "Pragmatic inferences modulate N400 during sentence comprehension: evidence from picture-sentence verification". *Neuroscience Letters* 534: 246-251. https://doi.org/10.1016/j.neulet.2012.11.044

Ivanova, Olga, and Bello Viruega, Iria. 2019. "Pragmatic processing in second language: what can focus operators tell us about cognitive performance in L2?". In *Cognitive Insights into Discourse Markers and Second Language Acquisition*, edited by Iria Bello Viruega, Carolina

Bernales, Maria Vittoria Calvi, and Elena Landone, 67-92. Oxford: Peter Lang.

Just, Marcel, Carpenter, Patricia, and Woolley, Jacqueline. 1982. "Paradigms and processes in reading comprehension". *Journal of Experimental Psychology: General* 111: 228-238. https://doi.org/10.1037/0096-3445.111.2.228

Kuznetsova, Alexandra, Brockhoff, Per, and Christensen, Rune. 2017. "lmerTest package: tests in linear mixed effects models". *Journal of Statistical Software* 82: 1-26. https://doi.org/10.18637/jss.v082.i13

Ladd, Robert. 1980. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Lauter, Miriam. 2013. *If SOME folks are wise, are others otherwise? The interaction of context and emphasis in online scalar implicature processing*. BA thesis, Yale University.

Luck, Steven. 2014. *An Introduction to the Event-Related Potential Technique*. Cambridge: MIT Press.

Luke, Steven. 2017. "Evaluating significance in linear mixed-effects models in R". *Behavior Research Methods* 49: 1494-1502. https://doi.org/10.3758/s13428-016-0809-y

Papadopoulou, Despina, Peristeri, Eleni, Plemenou, Evagilia, Marinis, Theodoros, and Tsimpli, Ianthi. 2015. "Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children". *Lingua* 155: 98-120. https://doi.org/10.1016/j.lingua.2014.09.006

Papadopoulou, Despina, Tsimpli, Ianthi, and Amvrazis, Nikos. 2013. "Self-paced listening". In *Research Methods in Second Language Psycholinguistics*, edited by Jill Jegerski and Bill VanPetten. New York: Routledge.

Politzer-Ahles, Stephen. 2020. "What can electrophysiology tell us about the cognitive processing of scalar implicatures?". *Language and Linguistics Compass* 14: 1-22. https://doi.org/10.1111/lnc3.12401

Politzer-Ahles, Stephen, and Husband, E. Matthew. 2018. "Eye movement evidence for context-sensitive derivation of scalar inferences". *Collabra: Psychology* 4: 3. https://doi.org/10.1525/collabra.100

R Core Team. 2016. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. https://www.R-project.org/.

Recanati, François. 1989. "The pragmatics of what is said". *Mind and Language* 4: 295-329. https://doi.org/10.1111/j.1468-0017.1989.tb00258.x

Snedeker, Jesse, and Trueswell, John. 2003. “Using prosody to avoid ambiguity: effects of speaker awareness and referential context”. *Journal of Memory and Language* 48: 103-130. https://doi.org/10.1016/S0749-596X(02)00519-3

Spychalska, Maria, Kontinen, Jarmo, and Werning, Markus. 2016. “Investigating scalar implicatures in a truth-value judgment task: evidence from event-related brain potentials”. *Language, Cognition and Neuroscience* 31: 817-840. https://doi.org/10.1080/23273798.2016.1161806

Wagers, Matthew, Borja, Manuel, and Chung, Sandra. 2015. “The real-time comprehension of WH-dependencies in a WH-agreement language”. *Language* 91, 109-144. https://www.jstor.org/stable/24672222

Wu, Fuyun, Kaiser, Elsi, and Anderson, Elaine. 2012. “Animacy effects in Chinese relative clause processing”. *Language and Cognitive Processes* 10: 1489-1524. https://doi.org/10.1080/01690965.2011.614423

Appendix: List of critical stimuli

The 24 critical items are shown here in the critical versions with "type contrast". For the complete list of stimuli (showing the other conditions of the critical items, as well as the filler and practice items and the comprehension questions for all items), see https://osf.io/8nvm7/.

1. 我去了 / **这**个 / 遛狗公园 / **连**一只狗 / 都没看到 / 更不用**说**/ 有人 / 了
I went to / this / dog park / even one dog / didn't see / let alone / people / SFP

2. **买**完**东**西后 / **连**一个硬**币** / 都不剩 / 更不用**说** / 有**纸币** /在**钱**包里 /
After I finished shopping / even one coin / not left / let alone / bills / in my wallet / SFP

3. 学校里 / **连**一个学生 / 都没来 / 更不用**说** / 有老**师** / 在走**动** / 了
At school / even one student / didn't come / let alone / teachers / walking around / SFP

4. 他 / 今天早上 / **连**一杯咖啡 / 都没喝 / 更不用**说** / 喝牛奶 / 了
He / this morning / even one cup of coffee / didn't drink / let alone / drank milk / SFP

5. 休息室里 / **连**一个球**员** / 都没有 / 更不用**说** / 有教**练** / 在走**动** / 了
In the locker room / even one athlete / wasn't there / let alone / coaches / walking around / SFP

6. 今天 / 猫咪咖啡**馆** / 装**电线** / **连**一只猫咪 / 都没有 / 更不用**说** / 有客人 / 在 / 了
Today / the cat café / installing wires / even one cat / not there / let alone / customers / there / SFP

7. 餐**厅**里 / **连**一个服**务员** / 都没有 / 更不用**说** / 有客人 / 在 / 了
In the cafeteria / even one server / not there / let alone / customers / there / SFP

8. 放暑假了 / 幼儿园里 / **连**一个小孩儿 / 都没有 / 更不用**说** / 有老**师** / 在 / 了
Summer break / at the kindergarten / even one kid / not there / let alone / teachers / there / SFP

9. 家里的 / 地上 / **连**一根**头发** / 都没有 / 更不用**说** / 有垃圾 / 了

At home / on the floor / even one strand of hair / not there / let alone / trash / SFP

10. 这个 / 废弃的 / 纪念馆 / 很荒凉 / 连一只苍蝇 / 都没有 / 更不用说 / 有人 / 在 / 了
This / abandoned / museum / very desolate / even one fly / not there / let alone / people / there / SFP

11. 大家 / 都去 / 救援了 / 医院里 / 连一个护士 / 都没有 / 更不用说 / 有医生 / 在应诊 / 了
Everyone / all went / to help [with the disaster] / at the hospital / even one nurse / not there / let alone / doctors / seeing patients / SFP

12. 铅笔盒里 / 连一只铅笔 / 都没有 / 更不用说 / 有橡皮 / 在内 / 了
In the pencil case / even one pencil / not there / let alone / erasers / inside / SFP

13. 广场的 / 地上 / 连一根烟头 / 都没有 / 更不用说 / 有垃圾 / 了
Plaza's / ground / even one cigarette but / not there / let alone / trash / SFP

14. 餐具区 / 连一双筷子 / 都没有 / 更不用说 / 有叉子 / 了
In the silverware area / even one pair of chopsticks / not there / let alone / forks / SFP

15. 这个时间 / 路上 / 连一台汽车 / 都没有 / 更不用说 / 有人 / 了
At this time / on the road / even one car / not there / let alone / people / SFP

16. 他们 / 刚搬进 / 新家 / 橱柜里 / 连一个锅子 / 都没有 / 更不用说 / 有碗 / 了
They / just moved into / new house / in the cupboard / even one pot / not there / let alone / bowls / SFP

17. 饭盒里 / 连一粒米 / 都不剩 / 更不用说 / 有菜渣 / 了
In the lunchbox / even one grain of rice / not left / let alone / bits of vegetables / SFP

18. 上课期间 / 儿童游乐园里 / 连一个小孩 / 都没有 / 更不用说 / 有家长 / 在里面 / 了
During class time / at the kids' playground / even one kid / not there / let alone / parents / there / SFP

19. 衣帽间里 / 连一件衣服 / 都没有 / 更不用说 / 有帽子 / 挂着 / 了
In the coatroom / even one item of clothing / not there / let alone / hats / hanging / SFP
20. 这个超市 / 人好多 / 连一个购物篮 / 都没有 / 更不用说 / 有购物车 / 了
This supermarket / many people / even one basket / not there / let alone / shopping carts / SFP

21. 邮箱里 / 连一封广告信 / 都没有 / 更不用说 / 有邮件 / 在内 / 了
In the mailbox / even one spam letter / not there / let alone / letters / inside / SFP

22. 停车场 / 连一辆单车 / 都没有 / 更不用说 / 有汽车 / 在内 / 了
The parking lot / even one bike / not there / let alone / cars / inside / SFP

23. 客厅 / 刚整修好 / 连一把椅子 / 都没有 / 更不用说 / 有沙发 / 摆在地上 / 了
Living room / just renovated / even one chair / not there / let alone / sofa / sitting on the floor / SFP

24. 鞋柜 / 才刚装好 / 连一双拖鞋 / 都没有 / 更不用说 / 有皮鞋 / 摆在里面 / 了
Shoe cabinet / just installed / even one pair of slippers / not there / let alone / leather shoes / sitting inside / SFP