Context effects in linguistic information processing: Decomposing the focus effect in reading*

Morwenna Hoeks\textsuperscript{1,*}, Maziar Toosarvandani\textsuperscript{1} & Amanda Rysling\textsuperscript{1}

\textsuperscript{1}University of California Santa Cruz

*Corresponding author

Author note

The data reported in this paper, as well as the analysis code and materials are made available via the Open Science Framework and can be accessed on \url{https://osf.io/k6tbw/?view_only=71d86431090046929d56f1ba94dcc38b}. The results of these experiments were presented at the 2021 CUNY Conference on Human Sentence Processing. We have no conflicts of interests to disclose.

Correspondence concerning this article should be addressed to Morwenna Hoeks, Department of Linguistics, University of California Santa Cruz, 1156 High Street, Santa Cruz CA, 95064. Email: mhoeks@ucsc.edu

\textsuperscript{*}We are grateful to Adrian Brasoveanu, Charles Clifton, Scott Fraundorf, Jesse Harris, Dean McHugh, Kelsey Sasaki, Shayne Sloggett, Nick Van Handel, and Matt Wagers for their feedback, as well as audiences at the UC Santa Cruz Graduate Research Symposium and CUNY 2021 for their comments and questions.
Abstract

Investigations of linguistic focus in reading have found mixed results: Some report a decrease in reading times on focused material \cite{MorrisFolk1998, BirchRayner2010}, while others report an increase \cite{BirchRayner1997, BenatarClifton2014, LowderGordon2015, SloggettRyslingStaub2019}, without an explanation for which pattern of results will be found with which focusing structures. We report the results of three studies using the Maze task \cite{ForsterGuerreraElliot2009}, in which newness was crossed with focus, and the context preceding a focused target manipulated whether or not contrastive alternatives to that target word were mentioned. A slowdown in response times on focused targets compared to non-focused targets was found even when the focus and its comparator were both given information, suggesting a focus cost that does not reduce to newness. When alternatives to foci were contextually mentioned, the slowdown on new foci was significantly reduced, suggesting that presenting information about alternatives aids the reading of foci. Controlling for newness versus focus and contextual mention of alternatives clarifies earlier reading results: previous reading work only found a focus speed-up after contextual mention of contrastive alternatives with no newness contrast between foci and baselines, and only found a slowdown in the absence of alternatives. We propose that differences in the processing profiles of material focused by different devices—question/answer pairs versus particles versus clefts—suggest that the different focus structures vary in the information they make available to and the specific demands they place on comprehenders, but the processing of focus is nonetheless generally effortful, in line with Lowder and Gordon’s \cite{LowderGordon2015} proposal.

Keywords: Linguistic focus, Givenness/Newness, Alternatives, Maze task, Reading

Perceivers must continually use evidence from context to determine where, when, and how greatly to allocate their processing resources. Human languages have systematized devices to highlight or background more versus less important information in discourse, and so language processing provides a case of the real-time deployment of expert knowledge in resource allocation, one that is common to nearly all humans. Studies that have investigated the processing of linguistic focus—sometimes characterized as the most “prominent” or “important” information in an utterance—in eyetracking while reading have found mixed results. Some have reported a decrease in reading times on focused material \cite{MorrisFolk1998, BirchRayner2010}, while others observed an increase in reading times \cite{BirchRayner1997, BenatarClifton2014}.
Explanations for the differences in these effects have acknowledged that focus is a complex construct and that its effects are likely to be modulated by many factors. For example, focus often covaries with informational newness; foci are often prosodically prominent; and, a word or phrase can be focused using a variety of different syntactic constructions, which each have their own specific properties and may require different resource allocations.

The most detailed proposal about the comprehension of focus in sentence processing is advanced by Benatar and Clifton (2014). They argue that, while many factors seem to cooccur with focus, the distinction that best captures the processing of focus is the dimension of newness versus givenness of information. Information is given if it is semantically entailed by what has come before in the discourse; all information that is not given is new (Schwarzschild, 1999). New information would be expected to require more effort to process than given information, because the new information has not already been integrated into a comprehender’s understanding of the world. This might explain the longer reading times on foci than non-foci that Benatar and Clifton and others found. If new information is always in focus, then any slow down on foci might be attributed to their status as discourse new.

Under Benatar and Clifton’s newness/givenness-based account of focus processing effects, the studies which instead found shorter reading times on focused words (Birch & Rayner, 2010; Morris & Folk, 1998) may have done so for a variety of reasons. Perhaps these shorter reading times were found because these studies compared foci that occurred later in sentences with baseline words that occurred earlier, and so natural oculomotor differences in reading the positions of words on a line caused foci to be read more quickly. Perhaps these shorter reading times were found because these studies provided greater contextual support for the focus-containing target sentence, in the form of a preceding context sentence. Or perhaps, these factors interacted with the syntactic devices that were used to focus words in these studies: all of them used clefts (It was . . . that . . . ), which introduce extra inferences into sentence comprehension. For these reasons, the best study under this account would hold target sentences constant, use questions to manipulate the location of focus as the part of an answer sentence that provided the asked-for information, so that the focus manipulation did not introduce extra inferencing during the processing of the focused word, and provide stable contextual support for all target sentences.
Benatar and Clifton’s studies accordingly compared new focused with given non-focused target words. Sentence position and inferencing requirements were equalized across conditions by employing question-answer pairs and holding target sentences constant within each item. In their first experiment, the answer to a preceding question provided the target sentence. The baseline condition consisted of completely given information, as in (1a), in which the preceding context question introduces both the existence of the characters and their possible relationship. In test conditions, longer reading times were then observed on target focused words, such as Natalie, when either the information about their relationships were new (here, the information that Kyle cared about Natalie) but the target word itself was not, as in (1b), or when the target word was entirely new, as in (1c). This held for first fixations, gaze durations, and total reading times.

(1) a. A: I’m confused, does Kyle care about Natalie?  
   B: Kyle cares about Natalie, but he doesn’t show it. 
   b. A: Natalie is confused, does Kyle care about someone?  
   B: Kyle cares about Natalie, but he doesn’t show it. 
   c. A: Isabella is confused, does Kyle care about someone? 
   B: Kyle cares about Natalie, but he doesn’t show it.

This pattern of longer reading times in (1b) and (1c), which contained new information pertaining to the focused word, than (1a), which contained only given information pertaining to the focused word, was interpreted as support for Benatar and Clifton’s account.

Support for the suggestion that sentence positional differences contributed to the findings of shorter reading time on foci was found by Lowder and Gordon (2015), who demonstrated that longer reading times on focused material generalize to syntactically-focused pseudoclefts (e.g., what the secretary typed was...). This construction permitted control of the sentence position of target words across conditions, something which had been lacking in the earlier studies that used simple clefts (it was the...) and found shorter reading times on foci (Birch & Rayner, 2010; Morris & Folk, 1998). A sample item from Lowder and Gordon’s study is shown in (2).

(2) a. What the secretary typed was the official memo about... 
   b. Yesterday the secretary typed the official memo about...
c. It was the secretary that typed the official **memo** about...

Lowder and Gordon found longer reading times on focused target words as compared to defocused target words (which were defocused, because another word in their sentences were overtly focused) in both gaze durations and regression-path durations. But they were further able to demonstrate that as target words became more focused—that is, comparing reading times in order across (2c), (2b), and (2a)—reading times increased. This increase with degree of focus supports the suggestion that variation in the effect sizes observed in previous work could also have been due in part to the differences in baseline conditions that were employed: studies that compared a focused word against a neutral baseline (Birch & Rayner, 1997) found longer reading times for foci as compared to non-foci, while studies that compared a focused word against a de-focused baseline (Morris & Folk, 1998) found faster reading times on foci.

Lowder and Gordon’s results clarified that the effect of syntactically-cued focus is one of longer reading times once sentence position is controlled. But differences in sentence position between foci and non-foci could not, by themselves, account for the difference between Birch and Rayner’s (1997) finding of longer reading times on foci than non-foci versus Birch and Rayner’s (2010) finding of shorter reading times on foci than non-foci. Birch and Rayner’s (1997) first experiment employed both new foci and new baselines, as shown in the sample item in (3), where *suburb* was the target word that was compared across conditions and in neither condition was it mentioned before or entailed by anything preceding it. Birch and Rayner (1997) found longer second-pass reading times on foci and greater probability of regression from foci than non-foci, even though *suburb* was newly mentioned in both conditions.

(3) a. It was the **suburb** that received the most damage from the ice storm. focus
   b. Workers in the **suburb** hurried to restore power after the ice storm. non-focus

This pattern of longer reading times contrasted with the findings of Birch and Rayner (2010), who used the same cleft focusing structure as Birch and Rayner (1997), but with a context sentence presented before each one, as in their item shown in (4).

(4) Context: The tenants at the complex were sick an tired of all the noise coming from #204.
a. It was the **landlady** who confronted the woman who lived there. focus

b. The **landlady** confronted the woman who lived there. non-focus

It thus seems that contextual support played a role in Birch and Rayner’s (2010) findings of shorter reading times on foci versus non-foci, but this property may not have held of Morris and Folk’s (1998) stimuli, which were not preceded by separate context sentences. An example item from Morris and Folk’s study is shown in (5), in which the target word *accountant* was compared across conditions. Notably, Morris and Folk’s conditions compared *accountant* in focus and de-focus, that is, with a different word from accountant in the syntactically focused position in (5b), unlike the simple declarative sentence with no special focus on the target word that served as a baseline for Birch and Rayner.

(5)  

a. While the waiter watched, it was the **accountant** who balanced the ledger a second time. focus

b. It was the waiter who watched while the **accountant** balanced the ledger a second time. defocus

It would be surprising if Morris and Folk’s items provided enough contextual support to facilitate the processing of their focused words, but Lowder and Gordon’s items did not. The pseudocleft structure employed by Lowder and Gordon likely provides more contextual support before its focus than the simple cleft; in the case of the sentence beginning *What the secretary typed was the official memo about...*, for example, a secretary typing was already introduced before the focused word *memo*, making that focused word much more expected in context than if it were early in the sentence and out-of-the-blue. Longer reading times on focused *memo* in Lowder and Gordon’s study would be mysterious under an account which appealed to contextual support to explain the shorter reading times on foci that previous studies had found. If newness is the primary driver of focus processing costs but it can be overcome by contextual support for focused words, Lowder and Gordon’s effects might be expected to be more like Birch and Rayner’s (2010) and Morris and Folk’s (1998) pattern of shorter reading times on clefted foci with contexts, instead of Birch and
Rayner’s (1997) pattern of longer reading times on clefted foci without contexts.

While Lowder and Gordon agreed with Benatar and Clifton that focus is a complex conjunction of many different properties, they suggested that focused material generally is more deeply encoded than non-focused material, with more effort expended to integrate it during language processing due to its greater importance in its sentences. This greater effort expended on focus would be expected to require more time, and would account for why focused material is advantaged in other tasks, for example, it is reliably better remembered than non-focused material (Birch & Garnsey, 1995; Gernsbacher & Jescheniak, 1995; McCoon, Ratcliff, Ward & Sproat, 1993; Singer, 1976).

In principle, Lowder and Gordon’s suggestion is not logically incompatible with the proposal that the processing profile of focus is generally due to a greater cost of processing new versus given material, even though their pattern of longer reading times on foci with greater contextual support when newness was a property of both foci and baselines is not straightforwardly explained under Benatar and Clifton’s account.

But, there is a test case which would separate Lowder and Gordon’s account of the focusing cost as due to general deeper processing from Benatar and Clifton’s account of the focus cost as largely due to the cost of new information. If focus effects are generally reducible to the givenness/newness distinction, then it would be unexpected to find longer reading times for focused material than non-focused material when the focus and its comparator non-focus were both given (and in the absence of confounding factors like different sentence positions between focus and comparator or the introduction of extra inferences by a specific focusing structure). In contrast, Lowder and Gordon’s proposal that focused material is simply more deeply encoded and effortfully processed than non-focused material would account for longer reading times on focused material even when both are given and all else is held equal.

We accordingly tested a full cross of givenness versus newness and focused versus non-focused targets in Experiment 1. It found that comprehenders take longer to respond to given foci than comparable given non-foci. Experiment 1 and subsequent experiments were initially designed to be implemented in eyetracking while reading, but due to the global Covid-19 pandemic, the Maze task (Forster et al., 2009) was employed as an alternative that could be safely run remotely. This seemed appropriate, because there is suggestive evidence that effects obtained using the Maze task are more similar to eyetracking while reading than the effects found in self-paced reading.
are (Witzel et al., 2012). While a close comparison with previous studies was lost, this change in method does allow conceptual replications of findings from the earlier literature to demonstrate the robustness of those effects across tasks, and so may increase our confidence in those effects’ generality beyond task-specific conditions.

**Experiment 1**

In Experiment 1, two factors — focus (focus vs. not) and newness (given vs. new) — were fully crossed, with focus modulated by preceding questions, which do not introduce the kind of extra inferences that clefts as focusing devices do. This provided a test of whether focus has an effect in the absence of newness, an outcome that would not be expected if the cost of focus processing is generally due to the cost of integrating new material. This was the first study to compare focus and baseline conditions that were both given information.

**Method**

**Participants**  51 participants were recruited via the Prolific platform for web-based research and were paid a $12 hourly rate for their participation. All participants were native speakers of English and gave explicit consent to participate. Participants that had an accuracy of less than 80% on the comprehension questions or that did not complete more than 70% of the Maze sentences were excluded from analysis. Data from 48 participants were included in the analysis; 3 participants were excluded because they failed to complete more than 70% of the Maze sentences.

**Materials** In all the experiments presented here, every item took the form of a short dialogue between two speakers, Speaker A and Speaker B. Speaker A first introduced a short premise, followed by a question. Speaker B’s utterance formed a response to the question from Speaker A. Speaker A’s utterance was considered the context sentence and was presented all at once on a single screen. Speaker B’s utterance was considered the target sentence and was presented using the Maze task. Within one item, the same sentence was the target for every condition, in order to ensure that differences across conditions would only be due to preceding context sentences. Within each target sentence, measurements on a single target word were expected to particularly reflect the effects of preceding contexts.
Preceding context questions determined whether a target word was new or given by either mentioning that target word in the question or not. Orthogonal to this manipulation of newness, preceding questions determined whether target words received narrow focus (NF) or broad focus (BF) by asking for differently specific information. Narrow focus questions were ones to which a following target word on its own would provide a complete answer; broad focus questions were ones to which a target word alone would not seem a complete answer. We employ the distinction between narrow and broad focus rather than the distinction between focused and neutral words, because focus in this study is manipulated by which parts of a target sentence provides the answer to a question. Our narrow versus broad focus conditions are analogous to Lowder and Gordon’s (2015) focused versus neutral conditions. Here, the narrow versus broad distinction captures Lowder and Gordon’s point that differences in degree of focus matters; the degrees of focus in the answers to questions can be understood as the proportion of the focus of a sentence that single word encompasses. In narrow focus conditions, the single focused word would be a complete answer to a preceding question and is the entirety of the focus. In broad focus conditions, the words that must be included in the focus are more numerous; in the case of these stimuli, they are the entirety of a phrase. This is illustrated in the example experimental item shown in (6) below. In (6), the target word is lawyer.

(6) **Speaker A:** This company often makes bad decisions, but...
   a. Did they hire a [lawyer] last fall, or an accountant?  
   b. Did they hire a [lawyer] last fall? 
   c. Did they hire an accountant last fall? 
   d. What did they announce last time?

**Speaker B:** I think they announced they hired a [lawyer] last fall, but I’m not sure.

In response to narrow focus questions, as in (6a) and (6c), lawyer would be a complete answer. Across all items, for creating narrow focus and givenness on the target word, alternative questions were used (i.e., questions in which two alternatives are given in the form of a disjunction). Since the answer to such a question is expected to be one of the mentioned alternatives, the answer was
either accountant or lawyer in the case of (6a). Therefore, the questions in the NF given conditions put only the target word lawyer in focus in the target sentence.

The NF new items always employed polar questions (i.e., questions whose expected answers are either confirmative or negative) that mentioned a different alternative from the one mentioned in the target sentence. The target sentence would therefore be unambiguously interpreted with corrective narrow focus on the target word.

After broad focus questions, lawyer in the target sentence would be part of a larger focused phrase, because a whole phrase from the target sentence would be required in order to provide a complete answer to the preceding questions. The BF given condition always used polar questions as well, but in these questions the alternative was the same as in the target sentence. This puts the target sentence as a whole in broad focus, as is the case in (6b). This had the result that both the wh-question in (6d) and the polar question in (6b) put at least the whole phrase they hired a lawyer last fall in focus, because this is the phrase that forms a congruent answer to each of these questions. Although it would be less natural, it is still technically possible that (6b) could be interpreted with narrow focus on the target word; there is nothing that prevents a reader from interpreting this as a narrowly focused phrase. However, evidence from interpretation and completion studies supports the assumption that comprehenders default to the broadest possible focus that is supported by the context (Harris & Carlson, 2014, 2017), and this accords with theoretical semantic assumptions as well. But, even assuming that narrow focus is more costly than broad focus, and that a narrow focus parse was maintained in at least some of our items’ broad focus given conditions, the estimated effect of focus from our study would be, if anything, slightly diminished and so decrease the likelihood that we would find an effect of focus in given conditions, because we expect narrow focus to be more costly to process than broad focus.

It is worth addressing a concern raised by an anonymous reviewer of an earlier version of this work: in order to manipulate newness/givenness in these stimuli, more material immediately before the target word was repeated in the given conditions (6a) and (6b) than in the new conditions (6c) and (6d). Any potential effect of the newness/givenness difference in these stimuli was thus perfectly confounded with repetition differences, with the result that expected shorter response times on given conditions relative to new conditions could be due to the simple repetition of more words across context and target in these; this was the problem that led Benatar and Clifton to
adopt hypernyms in their second experiment. This is not a concern for the present study, however, because the key prediction which Benatar and Clifton’s account cannot make is that an effect of focus can be found, after controlling for context, extra inferences, and syntactic structure, even when a focus and its baseline comparator are both given. The fact that so much more linguistic material is repeated across context and target sentences in the present given conditions does not obviate the givenness of those conditions here, and so these stimuli can serve as a test of this prediction. Importantly, even under the reasonable assumption that the repetition of lexical items in given conditions is the entirety of what underlies their expected shorter response times, this reducability would make it more difficult to observe the predicted effect of focus, because the target word would be primed, and responses to it would be facilitated, in both the narrow and broad focus given cases. This inherent confounding of the newness/givenness distinction with repetition in our items is thus not a confound for the effect that this study seeks to demonstrate: the presence of a focus effect even when both a focus and its non-focus comparator are given, and so the conceptual separability of focus from the givenness/newness distinction.

In each item, the target word was always followed by an adverbial phrase (last fall) which served as a spillover region. This spillover region was also always followed by a second clause (but I’m not sure), to ensure naturalness of the target sentence in the BF given condition.

In total, 48 items were constructed, each with the four conditions illustrated in (6). All items for Experiment 1 can be found in Appendix II; these were normed in an acceptability judgment study, the results of which can be found in Appendix I. Another 96 filler items which also consisted of multi-line discourses were interspersed with test stimuli. Using a Latin Square design, all 48 items were counterbalanced over 4 lists, such that each participant saw one condition from every item.

**Procedure** The Maze task is similar to the more commonly used self-paced reading task in that response times are measured using button presses. But instead of simply pressing a button to advance to a following word each time a participant has read the current word, participants in the Maze task see each word in the target sentence presented alongside a distractor word (or foil). Participants must at every new word choose the correct continuation between the intended item and its foil, which would not make a sensible continuation.
Foils were automatically generated using the AutoMaze software developed by Boyce, Futrell, and Levy (2020). This algorithm selects distractor words that are of the same length as the target word, and that are predicted by NLP language models to have a poor fit to the given context. For each upcoming word, a conditional probability distribution is determined for potential foils of the same length given the preceding sentence context. The words with a predicted probability below a certain threshold (or, above a certain surprisal threshold) are then selected by the AutoMaze algorithm as the distractor. Word frequencies that form the input to these models are obtained from the Google Books Ngrams corpus (Michel et al., 2011).

An example of the AutoMaze output for one target sentence is given in (7) below. On the second line, the distractor word is presented below its corresponding word of the target sentence.

(7) I think Sarah said she wanted cake for dessert, but I am not sure. x-x-x goods Runes blue sum bottom knee sum classed, tax Sin far sat send.

In this way, sentences were presented incrementally, and the response time required to make and execute a decision about which word should continue a sentence was measured.

On every trial, participants first read a context sentence on one screen. On a subsequent screen, participants were presented with the start of the target sentence in the format of the Maze task. That is, only the utterance of Speaker B was presented incrementally; the utterance of Speaker A was presented all at once for normal reading. All experimental trials were followed by a comprehension question, which made sure that participants had read the context preceding the target sentence. This was because there was more cause for concern that participants might not read the contexts than that they might not read the target sentences. Participants had to read the beginning and all subsequent material of a target sentence in order to even make a decision about which word could form a potential continuation as the sentence went on. If they chose the wrong word in the Maze task, they were directed to the next item and their responses on the rest of the words in the target sentence were not recorded. But participants could successfully go through a whole target sentence in the Maze without having read its preceding question, and so comprehension questions were included after each trial that encouraged careful reading of the preceding context. For instance, the example item in (6) was followed by the comprehension question in (8).

(8) Is the company known for its strategic actions?
Before being presented with the target stimuli and fillers, participants read a short description of the task, followed by five practice items. Practice items were similar to experimental items in that they involved a short context sentence, followed by a sentence presented in Maze format and a comprehension question. After the short practice phase, the experimental items were presented along with the fillers in a pseudo-random order.

**Transparency & Openness**  All materials, data and analysis code of this and subsequent experiments are made available via the Open Science Framework and can be accessed at [https://osf.io/k6tbw/?view_only=71d86431090046929d56f1ba94dcc38b](https://osf.io/k6tbw/?view_only=71d86431090046929d56f1ba94dcc38b) This study’s design and its analysis were not pre-registered.

**Analysis**  Data were analyzed using R, version 3.6.3 ([R Core Team](https://www.r-project.org) 2021). Linear mixed effects models were fit using the *lme4* package (Bates & Sarkar, 2007). Models included fixed effects of focus and newness, with broad focus and given conditions treated as reference levels, and random slopes and intercepts for both subjects and items ([Baayen et al.](https://www.ua.ac.be/linguistics/baayen) 2008). Separate models were fit with log-transformed response times and untransformed response times as dependent measures. Due to the fact that multiple comparisons were performed on the same data, an absolute value of *t* of 2.29 (instead of 2.00) was considered to be the threshold for significance.

**Results**

Mean response times for the target word and its surrounding regions in all conditions are given in Table [1](#). They are plotted with 95% confidence intervals in Figure [1](#).

The mean comprehension question accuracy was 88%, and the mean completion rate of the maze target sentences of Experiment 1 was 87%.

Tables [2](#) and [3](#) present the fixed effects results for the models of Experiment 1 log-transformed response times and untransformed response times on target words, respectively. Both models found two significant effects. First, significantly positive estimates of focus indicate that targets in narrow focus were responded to more slowly than targets that were part of a broad focused phrase. Second, significantly positive estimates of newness indicate that responses were slower in the new conditions compared to the given conditions. The interaction estimates did not reach
Table 1: Mean RT and standard error of the mean in each condition two words before, at, and two words after the target word.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Previous -1</th>
<th>Previous</th>
<th>Critical region</th>
<th>Spillover</th>
<th>Spillover +1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF (alt), given</td>
<td>686.71 (9.33)</td>
<td>643.62 (7.87)</td>
<td>774.52 (11.62)</td>
<td>741.49 (10.10)</td>
<td>711.16 (11.63)</td>
</tr>
<tr>
<td>BF (no alt), given</td>
<td>720.57 (11.53)</td>
<td>664.96 (8.81)</td>
<td>724.32 (10.66)</td>
<td>736.01 (11.00)</td>
<td>701.30 (10.12)</td>
</tr>
<tr>
<td>NF (alt), new</td>
<td>732.08 (12.91)</td>
<td>663.40 (8.42)</td>
<td>952.54 (15.64)</td>
<td>790.33 (12.16)</td>
<td>718.12 (11.38)</td>
</tr>
<tr>
<td>BF (no alt), new</td>
<td>893.66 (17.01)</td>
<td>745.91 (11.08)</td>
<td>937.43 (15.83)</td>
<td>867.97 (15.97)</td>
<td>770.00 (13.22)</td>
</tr>
</tbody>
</table>

Pairwise comparisons revealed that narrow focused given target words were responded to more slowly than broad focused given target words, for both log-transformed ($t = 3.72$) and untransformed ($t = 2.92$) response times.

Table 2: Parameter values for fixed effects in mixed linear regression model of LogRTs in Experiment 1.

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.54506</td>
<td>0.02175</td>
<td>300.890</td>
</tr>
<tr>
<td>Focus</td>
<td>0.06801</td>
<td>0.01822</td>
<td>3.732</td>
</tr>
<tr>
<td>Newness</td>
<td>0.23841</td>
<td>0.03005</td>
<td>7.933</td>
</tr>
<tr>
<td>Focus:Newness</td>
<td>-0.04306</td>
<td>0.02986</td>
<td>-1.442</td>
</tr>
</tbody>
</table>

Discussion

In addition to being affected by newness versus givenness, response times on the target word in Experiment 1 also depended on whether the target word was in narrow or broad focus. Comparing only targets words that were given, response times were longer when targets were narrowly focused than when they were only a part of a broadly focused phrase. This is unexpected if focus effects were primarily driven by the newness/givenness difference in the absence of either inferences demanded by syntactic constructions or the contextual support provided by preceding conditions.
material. The stimulus sentences of the present study held the target sentence identical across all conditions, and so no condition’s target sentence introduced meaning inferences that the others’ did not. All of the conditions were preceded by a question which contained much of the same linguistic material in the target sentence, thereby providing contextual support for the target word in narrow focus. Indeed, in the given narrow focus conditions, the preceding questions may have provided more contextual support than in the given broad focus conditions, because the narrow focusing questions also provided an alternative to the target word. Nonetheless, participants took longer to respond to the given narrow focused target word than the given broad focused target word, exactly the opposite of the pattern that an account appealing to a combination of newness and contextual support to explain previous findings would have predicted. Instead, the difference between the broad versus narrow given conditions found here conceptually replicates Lowder and Gordon’s observation that words take longer to read as they become more focused once sentence position is held constant, but at present with a different syntactic structure and task.

Figure 1: Experiment 1: mean RT in each region in each condition. Error bars represent the 95% confidence interval.
<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>725.07</td>
<td>18.13</td>
<td>39.994</td>
</tr>
<tr>
<td>Focus</td>
<td>50.36</td>
<td>17.05</td>
<td>2.954</td>
</tr>
<tr>
<td>Newness</td>
<td>210.88</td>
<td>30.63</td>
<td>6.886</td>
</tr>
<tr>
<td>Focus:Newness</td>
<td>-32.21</td>
<td>30.49</td>
<td>-1.057</td>
</tr>
</tbody>
</table>

Table 3: Parameter values for fixed effects in mixed linear regression model of raw RTs in Experiment 1.

**Interim discussion**

Experiment 1 was the first study to manipulate newness versus givenness and narrow versus broad focus independently of each other and to test for an effect of focus within entirely given material. In the process, the narrow versus broad distinction was achieved by the inclusion or exclusion of contextual alternatives, e.g., accountant, to the ultimate target word, e.g., lawyer.

The role of contextual alternative expressions in focus processing has been little explored in eyetracking studies of the reading focused words themselves (studies that have examined the role of focus in the subsequent processing of alternatives are discussed after the presentation of novel results). In this section, we present a review of the previous eyetracking while reading literature on focus processing, which suggests that the role of alternatives in the processing of focus is one of those as-yet unaccounted for factors referred to by Lowder and Gordon and Benatar and Clifton that will be crucial for the construction of a complete understanding of focus processing.

The alternatives to a focus are all those linguistic expressions which could have coherently taken its place in a sentence (Rooth, 1985). For example, in a sentence like (9), the alternatives to cake include, amongst other things, steak and cookies, since both expressions could be substituted for it to form a grammatical sentence. But, because steak is rarely eaten for dessert, comprehenders would generally not find it to be a relevant alternative. In other words, it is not a contextual alternative to cake for this sentence in most contexts.

(9) It was cake that Sarah wanted for dessert.

An alternative expression such as cookie, however, would likely be highly relevant, and therefore
would count as a contextual alternative, unless the comprehender of (9) also had access to some situational knowledge that ruled cookies out for other reasons.

Contextual alternatives factor into the inferences that clefts and focus particles like *only* give rise to. In (9), the comprehender is likely to understand that Sarah did *not* want cookies for dessert. Whether they draw this inference or not depends on their world knowledge, the content of the sentence itself, and the information provided by the preceding discourse context. Notably, question-answer pairs also give rise to a similar inference. In (10), the answer implies that Sarah did not want anything else, besides cake for dessert, including possibly cookies.

(10)  
  a. What did Sarah want for dessert?  
  b. Sarah wanted cake for dessert.

The strength of this inference in simple question-answer pairs is weaker, with comprehenders less likely to draw it in every context ([Hintikka, 1976] [Groenendijk & Stokhof, 1984] [van Rooy, 2003]).

Non-reading psycholinguistic studies employing priming, lexical decision, and memory tasks have shown that, when a focused expression is encountered, linguistic expressions that contrast with that focus become more strongly activated compared to expressions that are non-contextually semantically associated with the focus (e.g., [Braun & Tagliapietra, 2010] [Fraundorf et al., 2010, 2013] [Gotzner et al., 2016] i.a.). These studies strongly suggest that integrating a sentence’s meaning requires comprehenders to not only represent what that sentence described, but also to calculate alternatives to what was asserted. These non-reading results accord well with the theoretical semantic literature in linguistics. In the standard linguistic theory, foci are the word(s) in a sentence that must be contrasted with alternative expressions in order to understand what the sentence means ([Rooth, 1985, 1992]). A survey of the eyetracking while reading literature in terms of the alternatives to foci that were present or absent in stimulus sentences reveals that this dimension of focus perfectly demarcates studies by their reading time patterns.

All of the studies which report decreased reading times on focused material are ones in which potential alternatives to focused words were presented before those foci. This was true of Birch and Rayner (2010), whose stimuli are repeated in (11).

(11)  
  Context: The tenants at the complex were sick and tired of all the noise coming from #204.
a. It was the **landlady** who confronted the woman who lived there. focus
b. The **landlady** confronted the woman who lived there. non-focus

What (11a) conveys is that it was the landlady and not one of the tenants who confronted the woman who lived in the noise-making apartment. Crucially for the present discussion, the word *tenants* both can serve as an alternative expression to focused *landlady* and was presented in the preceding context sentence; this was systematic throughout Birch and Rayner’s (2010) items. This property also held of Morris and Folk’s (1998) stimuli.

In contrast, the eyetracking while reading studies which report reading time slow-downs on focus did not present alternatives to the focused material in preceding context sentences. This was true of Birch and Rayner’s (1997) first experiment, which did not employ preceding context sentences, and so did not present alternatives to target foci. This property was also true of Lowder and Gordon’s (2015) stimuli, even though these provided general contextual support for the target word. This lack of contextual alternatives to target foci also held for previous studies employing question-answer pairs as focusing devices (Birch & Rayner, 1997, experiments 2 and 3; Sloggett, Rysling, & Staub, 2019). Even in Benatar and Clifton’s (2014) studies, where stimuli were presented with preceding discourse context, the target words were difficult or impossible to understand as having contrastive alternatives. An example item from their first study is repeated in (12).

(12) a. A: I’m confused, does Kyle care about Natalie?
   B: Kyle cares about **Natalie**, but he doesn’t show it. given
b. A: Natalie is confused, does Kyle care about someone?
   B: Kyle cares about **Natalie**, but he doesn’t show it. new, rep
c. A: Isabella is confused, does Kyle care about someone?
   B: Kyle cares about **Natalie**, but he doesn’t show it. new, no rep

Because Benatar and Clifton used proper names in their new no repetition condition, which lack descriptive context. For a comprehender of (12) to establish that *Isabella* was a contextual alternative to a person named *Natalie*, either more contextual support or additional world knowledge would have been necessary. This is an easy intuition to grasp, when it is compared with the results from lexical priming for contrastive alternatives to foci (e.g., Braun & Tagliapietra, 2010; Fraun-
Inhibition Facilitation Construction Alternatives Newness

<table>
<thead>
<tr>
<th>Study</th>
<th>Inhibition</th>
<th>Facilitation</th>
<th>Construction</th>
<th>Alternatives</th>
<th>Newness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birch &amp; Rayner (2010)</td>
<td>×</td>
<td>✓</td>
<td>clefts</td>
<td>present</td>
<td>new new</td>
</tr>
<tr>
<td>Morris &amp; Folk (1998)</td>
<td>×</td>
<td>✓</td>
<td>clefts</td>
<td>present</td>
<td>new new</td>
</tr>
<tr>
<td>Ward &amp; Sturt (2007)</td>
<td>×</td>
<td>×</td>
<td>wh-phrase</td>
<td>absent</td>
<td>new new</td>
</tr>
<tr>
<td>Birch &amp; Rayner (1997), Exp. 1</td>
<td>✓</td>
<td>×</td>
<td>clefts</td>
<td>absent</td>
<td>new new</td>
</tr>
<tr>
<td>Lowder &amp; Gordon (2015)</td>
<td>✓</td>
<td>×</td>
<td>(pseudo)clefts</td>
<td>absent</td>
<td>new new</td>
</tr>
<tr>
<td>Birch &amp; Rayner (1997), Exp. 2</td>
<td>✓</td>
<td>×</td>
<td>q/a pairs</td>
<td>absent</td>
<td>new given</td>
</tr>
<tr>
<td>Benatar &amp; Clifton (2014), Exp. 1, 2</td>
<td>✓</td>
<td>×</td>
<td>q/a pairs</td>
<td>absent</td>
<td>new given</td>
</tr>
<tr>
<td>Benatar &amp; Clifton (2014), Exp. 3</td>
<td>✓</td>
<td>×</td>
<td>indefinites</td>
<td>absent</td>
<td>new given</td>
</tr>
<tr>
<td>Sloggett et al. (2019)</td>
<td>✓</td>
<td>×</td>
<td>q/a pairs</td>
<td>absent</td>
<td>new given</td>
</tr>
</tbody>
</table>

Table 4: Overview of previous investigations of focus in reading. Inhibition versus Facilitation columns indicate whether a study reported slower or faster reading times on focused material, respectively. Construction indicates which focusing structure was employed. Alternatives indicates whether alternative expressions to target words were presented. Newness indicates the new versus given status of the focus condition versus the condition(s) to which it was compared.

dorf et al., 2010; 2013; i.a.): the name Natalie would not be expected to prime the name Isabella. In their second study, Benatar and Clifton used hypernyms for target words, which contrasted with hyponyms in the preceding context. However, these expressions would not even count as alternatives in the first place, as alternatives must be exclusive (consider the infelicity of #I own a poodle, but not a dog). Thus, Benatar and Clifton’s studies, too, fit the pattern across the eyetracking while reading literature: faster reading times on foci which were presented after contrastive alternatives, but slower reading times on foci in the absence of (unambiguous) contrastive alternatives.

The results of the entire literature are summarized in Table 4. It is only the difference between the presence versus absence of alternatives to foci that demarcates the faster from the slower reading of foci across this earlier literature. We take this as suggestive evidence that the alternatives-based understanding of focus that is employed in the theoretical linguistic literature may be useful for building theories of language processing as well.
Upon inspection of Table 4, several other patterns are apparent. All reading studies of focus before the present Experiment 1 had tested only new foci. All reading studies that had employed question-answer pairs as focusing mechanisms did not presented alternatives to foci. None of these studies investigated the reading of foci that were marked by focus-sensitive particles, such as *only*, which obligatorily focus an element in their scope. And, while the closest comparisons in the literature so far were between studies that employed clefts or (pseudo)clefts with or without the presence of possible alternatives to the foci, there still is not a minimal comparison of clefts that are always preceded by contexts, which themselves differ only in whether alternatives are present versus absent. This is because Birch and Rayner’s two cleft studies differed in the presence versus absence of entire contexts and Lowder and Gordon’s pseudoclefts differed from Morris and Folk’s clefts in sentential positions as well as not having contexts.

As other authors have noted, each focusing device carries with it certain unique demands, and Table 4 shows that the speed-up in reading times on focused material after the presentation of contrastive alternatives has only been demonstrated with clefted foci. It is possible that this pattern would hold only of clefts, or would only hold of structures that shared some property with clefts; this would account for why it was not observed in the present Experiment 1.

Support for the hypothesis that the effect of contextual alternatives interacts with the differences among focus constructions comes from the theoretical semantics literature, which distinguishes between *free* versus *associated* foci (Jackendoff, 1972; Rooth, 1985). Associated foci are those that are signaled by a particular syntactic construction, such as (pseudo)clefts or focus particles (e.g., *only, even*), while free foci are those that are merely mandated by context, such as by a preceding question, not by any expression in their immediate sentence. Not only do clefts and focus particles generally cause comprehenders to calculate more construction-specific inferences or presuppositions than free foci, but also the foci associated with these constructions are more predictable in incremental processing, because many of these devices’ foci must come after them. The locations of the foci of clefts are predictable with a high degree of certainty before those foci have themselves been fixated, because it is always a word very soon after *was* (or *is*) that is focused. Similarly, the foci of clefts and *only* always follow, although this can be at a small distance. In contrast, a comprehender can only predict the positions of free foci that are later in the sentences in which they occur, because it is only after some linguistic material has elapsed that
this material could have sufficiently narrowed the possible continuations that would be congruent with the preceding focusing structure.

This point is illustrated by the example sentences in (13), in which a comprehender does not know when they first turn to reading the answer to (13a) whether the response will be more like (13b), where the answer and so the focus is later in the sentence and by that point more predictable, or more like (13c), where the answer and focus is unsignalled in the sentence before it occurs.

(13)  
   a. What did Sarah want for dessert?  
   b. Sarah wanted **cake** for dessert. late focus  
   c. **Cake**, I think it was. early focus

It is thus the case that different focusing constructions could have different processing profiles in the presence or absence of earlier contrastive alternatives, not because of differences in inferences or presuppositions that these constructions trigger, but instead because of differences in how readily a reader can predict that a focus will occur. In line with Lowder and Gordon’s proposal of deeper encoding and greater integration of focused material, it may be that readers are better able to allocate their resources toward an upcoming focus ahead of time when reading an associated focus than a free focus. Combining such an account with an alternatives-based understanding of focus, it may be that contrastive alternatives can more greatly facilitate the processing of associated than free foci, because the amount of effort required to identify a focus is already eased in the processing of associated foci, and so it is in these constructions that readers are then more quickly able to take advantage of the pre-activation of contrastive alternatives to the foci they will need to comprehend. In the case of free foci, however, comprehenders must identify that material has been focused by preceding context anew in each situation; there are not such clear and consistent cues as dedicated lexical items like *only* or *it was a...* to signal focus.

In light of this similarity between focus particles and clefts, these constructions were tested, respectively in the presence or absence of alternatives in Experiment 2 and Experiment 3, which were designed to reveal the interaction of contrastive alternatives with associated focus constructions.
Experiment 2

The only difference between Experiment 1 and Experiment 2 is that in the latter, the focus particle *only* was added to the target sentences. This particle was unambiguously associated with the target word and served two purposes. First, it provided a cross-experiment comparison between free and associated focus, by making the stimuli of Experiment 1 and Experiment 2 differ in their focusing constructions and so how clearly signaled the position of foci were. Second, the focus particle put the target word into narrow focus in all of the conditions of Experiment 2. It therefore allowed assessment of the effect of providing alternatives to target words in preceding contexts when focus status is held constant.

Method

Participants 58 native speakers of English were recruited via Prolific. All participants were compensated at a $12 hourly rate. Completion of the experiment usually took 50 minutes including the practice phase. Data from 48 participants were included in the analysis; 10 participants were excluded because they failed to complete more than 70% of the Maze sentences.

Materials The stimuli of Experiment 2 were exactly the same as in Experiment 1, except for the presence of the particle *only* immediately before the target word in each the target sentence. An example of an item is shown in (14) below.

(14) **Speaker A:** This company often makes bad decisions, but...

a. Did they hire a [lawyer] last fall, or an [accountant]? (NF) alt, given

b. Did they hire a [lawyer] last fall? (NF) no alt, given

c. Did they hire an [accountant] last fall? (NF) alt, new

d. What did they announce last time? (NF) no alt, new

**Speaker B:** I think they announced they hired only a [lawyer] last fall, but I’m not sure.

As in Experiment 1, a newness manipulation determined whether the target word was new or given by the time participants read it. Crucially differently from Experiment 1, the target word of
every condition in Experiment 2 was always narrow focused. In (14), lawyer is associated with only, which puts it in narrow focus. This position of only immediately before the target word prevented it from being interpreted as associated with any other word or phrase in the sentence. In order to facilitate comparison of these conditions with the ones from Experiment 1, the label “NF” is shown next to all the conditions of Experiment 2 in (14).

In order to interpret the meaning of only in a target sentence, comprehenders require contextually relevant alternatives to the target word, because the meaning of only is that nothing other than its associated focus is true in that context. As noted above, the preceding context questions used in both Experiment 1 and Experiment 2 manipulated the presence versus absence of an alternative to the target word. In Experiment 2, where all target words are focused by their association to only, this manipulation thus assessed the effect of explicitly provided alternatives in processing focus. Since an alternative question like that in (14a) presupposes that the mentioned alternatives are the only possible hires, accountant formed a salient alternative to the target lawyer. Similarly, the polar question in (14c) explicitly mentioned an alternative to the target word, accountant, but did not mention the target word itself. Thus, (14a) and (14c) are labeled “alt,” while (14b) and (14d) are labeled “no alt.” All materials of Experiment 2 were assessed in an acceptability judgment study, the results which can be found in Appendix I.

**Procedure** As in Experiment 1, the target sentences in Experiment 2 were implemented in the Maze task, in which response times were measured as the time it took for participants to choose between the actual continuation word and a foil.

Since the materials for Experiment 1 and Experiment 2 were the same except for the word only in the target sentence, the foils generated for Experiment 1 were used to create the foils for Experiment 2. To do so, the target sentences of Experiment 2 were used as the input to the AutoMaze algorithm to generate the appropriate foils for the word only in each item. Then, these foils for only were inserted into the foils that were already generated for Experiment 1. In this way, the differences between Experiment 1 and Experiment 2 were kept as minimal as possible, to ensure maximal comparability between the two experiments. Fillers, practice items, comprehension questions, and presentation lists were the same as in Experiment 1.
### Table 5: Experiment 2: mean RT and standard error of the mean in each condition two words before, at, and two words after the target word.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Previous -1</th>
<th>Previous</th>
<th>Critical region</th>
<th>Spillover</th>
<th>Spillover +1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NF) alt, given</td>
<td>776.95 (17.49)</td>
<td>676.20 (10.40)</td>
<td>799.06 (11.58)</td>
<td>758.62 (10.83)</td>
<td>735.89 (17.40)</td>
</tr>
<tr>
<td>(NF) no alt, given</td>
<td>793.35 (16.33)</td>
<td>683.25 (9.66)</td>
<td>791.93 (14.13)</td>
<td>782.73 (13.58)</td>
<td>747.24 (13.06)</td>
</tr>
<tr>
<td>(NF) alt, new</td>
<td>811.27 (16.95)</td>
<td>711.83 (12.45)</td>
<td>901.01 (13.32)</td>
<td>793.31 (12.19)</td>
<td>713.45 (11.16)</td>
</tr>
<tr>
<td>(NF) no alt, new</td>
<td>909.15 (17.25)</td>
<td>772.62 (16.40)</td>
<td>968.70 (17.32)</td>
<td>876.82 (15.27)</td>
<td>835.61 (18.91)</td>
</tr>
</tbody>
</table>

**Analysis**  
The analysis was the same as that of Experiment 1, except that fixed effects were newness and the presence versus absence of alternatives. The absence of an alternative to the focused target word was treated as the reference level of this factor.

**Results**

Mean response times for the target word and its surrounding regions in all conditions are given in Table 5. They are plotted with 95% confidence intervals in Figure 2.

The mean comprehension question accuracy was 88%, and the mean completion rate of the maze target sentences of Experiment 2 was 87%.

Fixed effects estimates for the model fitted to log-transformed responses are reported in Table 6; those for the model fitted to untransformed responses are reported in Table 7.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>6.64594</td>
<td>0.02132</td>
</tr>
<tr>
<td>Alternative</td>
<td>-0.02849</td>
<td>0.02232</td>
</tr>
<tr>
<td>Newness</td>
<td>0.10723</td>
<td>0.02257</td>
</tr>
<tr>
<td>Alternative:Newness</td>
<td>0.09016</td>
<td>0.02820</td>
</tr>
</tbody>
</table>

Table 6: Parameter values for fixed effects in mixed linear regression model of LogRTs in Experiment 2
As in Experiment 1, the significant positive estimates for newness indicate that new targets were responded to more slowly than targets that were mentioned in the preceding question. A significant interaction between newness and the presence of alternatives indicated that the difference in response times between the two new conditions was larger than the difference between the two given conditions. Pairwise comparisons on untransformed response times confirm that the effect of presence of alternatives does not reach significance in the given conditions ($t = .31$), while the difference between the new conditions with and without previously mentioned alternatives was significant ($t = −3.60$). Responses were thus slower in conditions without an alternative compared to those with an alternative only when the target was also new.

**Discussion**

Experiment 2 replicated the effect of newness found in Experiment 1: responses were slower when a target word was new compared to when it was given. This finding also replicated the results of Benatar and Clifton’s eye-tracking while reading studies in which new information focus was found to cause significant slowdowns.
Table 7: Parameter values for fixed effects in mixed linear regression model of untransformed RTs in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>803.153</td>
<td>19.503</td>
<td>41.180</td>
</tr>
<tr>
<td>Alternative</td>
<td>-3.699</td>
<td>24.699</td>
<td>-0.150</td>
</tr>
<tr>
<td>Newness</td>
<td>94.717</td>
<td>22.865</td>
<td>4.142</td>
</tr>
<tr>
<td>Alternative:Newness</td>
<td>86.692</td>
<td>34.821</td>
<td>2.490</td>
</tr>
</tbody>
</table>

Experiment 2 also found limited evidence that preceding contextual information modulates the reading of foci. Narrow foci that were preceded by a contextually-mentioned alternative expression to the target word were read faster than narrow foci that were not preceded by an alternative, but only when the target itself was new. Two possible reasons for this are that previously encountering the exact expression in focus or previously encountering an alternative to the focus aids in comprehending the focus itself. But it is also possible that, across the two new conditions, the alternative-mentioned condition contained more repetition of words before the target word than the no-alternative-mentioned condition. This could have facilitated processing of the alternative-mentioned condition throughout the sentence, as suggested by the generally lower reading times before and after the target word in this condition relative to the new one without an alternative. Without a difference between the given conditions, it is impossible to adjudicate between a contrastive alternatives-based versus simple repetition-based explanation for the faster response times to new alternative-mentioned conditions here in this study alone. However, the results of Experiment 3 on another kind of associated focus, clefts, may suggest that the lack of difference between the given conditions in Experiment 2 is, itself, a floor effect.

**Experiment 3**

The crucial difference between Experiment 2 and Experiment 3 is that in the latter, an *it*-cleft was used to focus target words, instead of the focus particle *only*. 
**Method**

**Participants** 53 native speakers of English were recruited via Prolific. Data from 48 participants were included in the analysis; 5 participants were excluded because they failed to complete more than 70% of the Maze sentences.

**Materials** The items of Experiment 3 consisted of modified versions of those of Experiment 2. An example of an item is in (15), below.

(15) **Speaker A:** This company often makes bad decisions, but...

   a. Did they hire a [lawyer] last fall, or an [accountant]? (NF) alt, given
   b. Did they hire a [lawyer] last fall? (NF) no alt, given
   c. Did they hire an [accountant] last fall? (NF) alt, new
   d. What did they announce last time? (NF) no alt, new

   **Speaker B:** I think they announced it was a [lawyer] that they hired, but I’m not sure.

As in Experiment 2, the preceding context questions of Experiment 3 manipulated whether an alternative to the expression in focus was either mentioned or not (alt vs. no alt) and whether the focus itself was previously mentioned or not (given vs. new). Like the focus particle *only*, the cleft structure (*it was a...*) caused all target words in all conditions of Experiment 3 to be unambiguously narrow focused.

Besides replacing *only* with a cleft, another difference in the target sentences between Experiment 3 and Experiment 2 was that the phrase that previously functioned as a short spill-over region (*last fall* in Experiment 1 and Experiment 2) was removed from the target sentence in Experiment 3 to make the target sentence slightly shorter and more natural. In addition, the verbs of which target words were direct objects were moved to immediately after target words, as in *lawyer that they hired* in (15). For this reason, Experiment 3 no longer confounded givenness with the simple repetition of the words immediately before the target word; if anything, it was the new condition without alternatives mentioned in (15d) that contained the most repetition across context and target sentences before the target word.
All materials of Experiment 3 were first assessed in an acceptability judgment study. The results of this norming study can be found in Appendix I. Fillers, practice items, and comprehension questions were the same as in the previous two studies.

**Procedure**  As in Experiment 2 and Experiment 1, target sentences were implemented in the Maze task. Maze foils for Experiment 3 were independently generated using the AutoMaze algorithm, with the result that the foils in this experiment were not directly based on those generated for Experiment 1 and Experiment 2. This was necessary, because the target sentences in Experiment 3 are of a different structure from the target sentences in Experiment 1 and Experiment 2. For this same reason, a direct comparison between response times obtained in these experiments and those from Experiment 3 would not have been possible regardless of the way in which the foils were generated.

**Analysis**  The analysis was the same as that of Experiment 2.

**Results**  Mean response times for the target word and its surrounding regions in all conditions are presented in Table 8. They are plotted with 95% confidence intervals in Figure 3.

The mean comprehension question accuracy was 86%, and the mean completion rate of the maze target sentences of Experiment 3 was 83%.

Fixed effects estimates for the model fitted to log-transformed response times are reported in Table 9; those for the model fitted to untransformed response times are reported in Table 10. Significantly positive estimates for newness again indicated a slowdown on new targets compared to targets that were mentioned in the previous question. Unlike Experiment 2, however, the models fitted to Experiment 3 revealed a small but significant effect of the presence of alternatives, indicating that foci were read faster in the presence of a contextual alternative than in the absence of one. Finally, a significant interaction between newness and the presence of alternatives was also found, suggesting that the difference in response times between the two new conditions was larger than the difference between the two given conditions. As in Experiment 2, pairwise comparisons on untransformed response times revealed that the effect of the presence of alternatives only reaches
<table>
<thead>
<tr>
<th>Condition</th>
<th>Previous -1</th>
<th>Previous</th>
<th>Critical region</th>
<th>Spillover</th>
<th>Spillover +1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NF) alt, given</td>
<td>691.65 (10.52)</td>
<td>681.21 (9.19)</td>
<td>823.68 (9.19)</td>
<td>680.58 (8.42)</td>
<td>670.26 (11.46)</td>
</tr>
<tr>
<td>(NF) no alt, given</td>
<td>697.15 (10.78)</td>
<td>693.09 (9.84)</td>
<td>877.81 (16.87)</td>
<td>683.79 (9.84)</td>
<td>678.52 (11.91)</td>
</tr>
<tr>
<td>(NF) alt, new</td>
<td>701.00 (11.80)</td>
<td>707.76 (10.44)</td>
<td>977.80 (16.66)</td>
<td>698.56 (9.43)</td>
<td>673.30 (10.68)</td>
</tr>
<tr>
<td>(NF) no alt, new</td>
<td>710.78 (11.44)</td>
<td>724.57 (10.59)</td>
<td>1160.42 (22.56)</td>
<td>736.85 (9.96)</td>
<td>736.54 (12.81)</td>
</tr>
</tbody>
</table>

Table 8: Experiment 3: mean RT and standard error of the mean in each condition two words before, at, and two words after the target word.

Table 9: Parameter values for fixed effects in mixed linear regression model of LogRTs in Experiment 3.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>6.65769</td>
<td>0.03370</td>
</tr>
<tr>
<td>Alternative</td>
<td>0.04607</td>
<td>0.02272</td>
</tr>
<tr>
<td>Newness</td>
<td>0.17234</td>
<td>0.02103</td>
</tr>
<tr>
<td>Alternative:Newness</td>
<td>0.11022</td>
<td>0.03919</td>
</tr>
</tbody>
</table>

Discussion

Experiment 3 replicated the newness effects reported above for both Experiment 1 and Experiment 2: In all experiments, responses were slower to new foci than given foci. Experiment 3 also replicated Experiment 2 in finding that the slowdown for new foci was smaller when a context mentioned an alternative expression to the target word. This effect from Experiment 2 was replicated even in the conditions of Experiment 3, in which it was the new conditions without preceding alternatives mentioned that contained the least repetition across contexts and targets. This suggests that, for associated foci generally, either previously encountering the expression in focus or previously encountering an alternative to the focus significantly helps in comprehending the focus.
General discussion

Experiments 1-3 demonstrate that focus effects in reading are not reducible to the newness/givenness distinction; instead, the appropriate understanding of focus for language processing research is with reference to the alternative expressions that a sentence’s material is most relevantly contrasted with. The results of Experiment 1 support this conclusion, because longer response times were found on narrow focused words compared to words that were part of broad focused phrases even after controlling for newness versus givenness. At the same time, Experiment 2 and Experiment 3 demonstrate that the earlier mention of alternative expressions can somewhat attenuate the cost of processing new material when that new material is also an associated focus. The results reported in this paper thus accord with the findings of Benatar and Clifton (2014), Birch and Rayner (1997), Lowder and Gordon (2015), and Sloggett et al. (2019), who all argued for a general processing cost of focus. But this general focus slowdown cannot be reduced to the newness/givenness distinction, even in the absence of extra inferences or contextual support. Our findings are more in
Table 10: Parameter values for fixed effects in mixed linear regression model of untransformed RTs in Experiment 3.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>826.99</td>
<td>31.93</td>
<td>25.901</td>
</tr>
<tr>
<td>Alternative</td>
<td>55.48</td>
<td>24.13</td>
<td>2.299</td>
</tr>
<tr>
<td>Newness</td>
<td>157.85</td>
<td>23.82</td>
<td>6.627</td>
</tr>
<tr>
<td>Alternative:Newness</td>
<td>128.39</td>
<td>44.78</td>
<td>2.867</td>
</tr>
</tbody>
</table>

accord with Lowder and Gordon’s interpretation of focus processing costs as deeper encoding or more effortful integration with greater degrees of focus.

The present findings are also potentially compatible with the speed-ups on focused material reported by Birch and Rayner (2010) and Morris and Folk (1998). Both Birch and Rayner (2010) and Morris and Folk made use of materials in which target words were new, alternatives to the expression in focus were explicitly mentioned, and target words were focused by clefts, just as in the present Experiment 3. While Experiment 3 only found that new foci which followed alternatives were read less slowly than new foci which did not follow alternatives, without evidence of a speed-up, the present studies’ baseline conditions were unlike Morris and Folk’s and Birch and Rayner’s. Looking more closely at Morris and Folk’s materials may provide a clue for an alternative explanation about why such a speed-up may have arisen. An example item is repeated in (16) below, in which an alternative, waiter, always preceded the target word accountant.

(16) a. While the waiter watched, it was the **accountant** who balanced the ledger a second time.
   
   b. It was the waiter who watched while the **accountant** balanced the ledger a second time.

It may be that this earlier alternative expression, waiter, was ultimately also understood as focused by the readers of these sentences, because it was understood in clear contrast with accountant. In other words, the target word accountant in the defocus condition might have received contrastive focus as well, because it seemed like a relevant alternative to waiter. In that case, the comparison made in this study would have been one between an ultimately contrastive focus without a pre-
ceding alternative as in (16b) and a focus inside a cleft with a preceding alternative as in (16a), where the latter type of focus gave rise to shorter reading times than the first. This could have been due to the focused target word in (16a) being more clearly demarcated as focused by its preceding cleft than the ultimately contrastively focused target word in (16b) requiring more inference on the comprehender’s part. If the speed-up in reading the focused target words in Morris and Folk’s study was due in large part to a combination of the presentation of alternatives before the target word and the clarity of focus marking provided by cleft constructions, then this explanation would extend to Birch and Rayner’s (2010) faster reading times on focused words as well.

If inherent ambiguity in the location and size of a free focus makes it more costly to process than foci associated with a cleft or only, which are overtly signaled, then the different patterns of processing times on foci after alternatives in the present Experiment 1 versus Experiment 2 and 3 could be understood as the construction-specific demands of focus processing. Morris and Folk’s and Birch and Rayner’s studies may, in fact, be better understood as more similar to comparing the broad focused, no-alternatives-mentioned, new condition of Experiment 1 as a baseline against the narrow focused, alternatives-mentioned, new condition of Experiment 3. At this point, it is not clear whether the facilitated reading of associated foci after contextual alternatives that we observe here is due to the particular properties of the syntactic expressions (clefts, only) they were associated with, or whether it is due to the general fact that they are associated at all (and hence, we do not know whether foci associated with other particles, such as too or even, would show the same effect).

One possible piece of evidence in support of the suggestion that free foci are generally more costly to process comes from self-paced reading studies reported by Fraundorf et al. (2013), who also showed slowdowns on foci that occurred even after the explicit mention of contextual alternatives. Unlike any of the studies discussed here thusfar, in Fraundorf et al.’s materials, foci were marked using font emphasis. No focusing device, whether contextual or syntactic, signalled the presence of the upcoming focus in advance of the emphasized word. Fraundorf et al.’s studies may therefore have yielded a slowdown even following contextually-mentioned alternatives, because their conditions had in common with free foci the property that comprehenders were not able to confidently anticipate a focus before they encountered it.

The studies reported here thus provide support for Benatar and Clifton’s and Lowder and Gor-
don’s suggestion that different focus constructions may all be processed slightly differently. The attenuation of a newness slowdown when alternatives to foci were explicitly mentioned was only observed for material that was focused by either the particle only or a cleft, that is, for associated foci. In Experiment 1, the narrow foci necessarily occurred in a context in which an explicit alternative was mentioned, but these new narrow-focused target words were not read faster than target words that were part of a new broad-focused phrase. Alternative expressions seem to be most useful when the focus structure of a sentence is clearly signalled.

If comprehending foci also requires some understanding of the most relevant contrastive alternatives to what a sentence conveys, then it would be expected that the contextual mention of alternatives can aid in the reading of foci. We thus propose that psycholinguistic theories adopt a conceptualization of focus in terms of contrastive alternatives in their reasoning at Marr’s (1982) computational level, i.e., reasoning about what the system’s goals are. At a more algorithmic level, alternative expressions may provide some semantic associate priming benefit to upcoming foci, and the process of fully comprehending a focus may encompass first the activation of semantically associated expressions, followed second by the narrowing of those associated expressions into only the set that would be contrastive in the current context, as has been suggested by Husband & Ferreira (2015), inter alia. This understanding of focus processing would accord well with both the studies that have found lexical priming of alternative sets from focused words (Braun & Tagliapietra 2010; Fraundorf et al. 2013, 2010; Gotzner et al. 2016; Husband & Ferreira 2015) and a growing body of reading studies that have demonstrated that comprehenders use the content of focused expressions to anticipate the upcoming mention of contrastive alternatives (Filik, Paterson & Liversedge, 2009; Ferreira & Lowder, 2016; Lowder, Ryan, Opie, & Kaminsky, 2021).

In sum, the present studies demonstrated an overall slowdown in response times for narrow foci compared to broad foci, extending to a new method the pattern found in the majority of earlier studies on focus in eyetracking while reading (Benatar & Clifton 2014; Birch & Rayner 1997; Lowder & Gordon 2015; Sloggett et al. 2019). But this focus slowdown cannot be reduced to the role of newness, because given narrow foci were responded to more slowly than given broad foci. Furthermore, substantial differences were found in the processing of different types of focusing devices; these need to be investigated in future research, but the minimal comparisons presented here suggest that considering the differential informativity of focusing devices will be a fruitful
avenue of research. For the present, this paper puts forward evidence that previous reading results
can be explained by adopting the appropriate (computational level) understanding of focus for
psycholinguistic theories. This is the same as the understanding of focus in formal linguistics—the
only property that unifies all focus constructions is the requirement that contrastive alternatives be
considered in order to understand the meaning of focused expressions. Such an alternatives-based
conceptualization of focus for language processing is compatible with results from eyetracking
while reading, the Maze task, self-paced reading, lexical priming, and event-related potentials,
which all converge on the conclusion that the comprehenders automatically consider alternatives
to focused expressions during the course of real time language processing.

References


Birch, S. L., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of
incremental processing difficulty. *Journal of Memory and Language, 111*, 104082.

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of


Hintikka, J. (1976). The semantics of questions and the questions of semantics: Case studies in the interrelations of logic, semantics, and syntax.


### 1 Appendix I: Offline acceptability ratings

The offline acceptability judgment studies discussed in this section aimed to establish the extent to which the materials used in the Maze task online reading studies were considered natural by native speakers of English. To that end, Experiments A.1-3 use the same stimulus and filler materials as Experiments 1-3. Since reduced acceptability ratings have been shown repeatedly to provide an indication of a significant processing cost, these offline studies also provided preliminary and convergent evidence for potential focus costs.

Participants were from the same population as Experiments 1-3 and recruited in the same way. Sentences were presented using the Ibex Farm platform for web-based experiments (Drummond, 2013).

In each trial, participants read a full dialogue on a single screen and were asked to judge the naturalness of the full discourse on a 4-point Likert scale. The practice items provided guided feedback to make sure participants were familiar with the use of the scale.

All of the studies reported here were analyzed with mixed effects ordinal regression models fitted to the rating data using the `clmm` function of the ordinal package in R (R Core Team, 2021; Christensen, 2019). All fixed and random effects structures parallel those used for the Maze studies, unless otherwise noted.
1.1 Experiment A.1

(17) **Speaker A:** This company often makes bad decisions, but...

a. Did they hire a lawyer last fall, or an accountant? NF (alt), given
b. Did they hire a lawyer last fall? BF (no alt), given
c. Did they hire an accountant last fall? NF (alt), new
d. What did they announce last time? BF (no alt), new

**Speaker B:** I think they announced they hired a lawyer last fall, but I’m not sure.

This acceptability rating study also aimed to establish whether, in the NF given condition as in (17a), the eventual target word was considered a natural alternative expression to the alternative mentioned the preceding question and vice versa. If the target and the alternative expression were indeed proper alternatives to each other, it would be expected that it would not matter which one was mentioned in the question and which one was mentioned in the target sentence. In Experiment A.1, both the intended question/answer pairs and the question/answer pair in which the position of the target and the alternative expression were switched were tested.

The identity of target and alternative expression was treated as a between-subjects manipulation: one group of participants (n=48) were presented with the set of items that were be used in our reading studies, while a second group of participants (n=48) were presented with the version of all the items that had the target and the contextual alternatives switched.

In a separate model, the identity of the target word, again with two levels, was added as a between-subjects fixed effect. A t-value of 2 will be considered to be the critical value for significance. The broad focus and given conditions were treated as baselines throughout.

1.2 Experiment A.2

(18) **Speaker A:** This company often makes bad decisions, but...

a. Did they hire a lawyer last fall, or an accountant? (NF) alt, given
b. Did they hire a lawyer last fall? (NF) no alt, given
c. Did they hire an accountant last fall? (NF) alt, new
Table 11: Experiment A.1: mean rating and standard error of the mean by condition and by target and contextual alternatives.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Target Identity</th>
<th>Mean</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF (alt), given</td>
<td>alt1</td>
<td>3.34 (0.043)</td>
<td>alt1</td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>3.30 (0.055)</td>
<td>alt2</td>
</tr>
<tr>
<td>BF (no alt), given</td>
<td>alt1</td>
<td>3.29 (0.035)</td>
<td>alt1</td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>3.27 (0.050)</td>
<td>alt2</td>
</tr>
<tr>
<td>NF (alt), new</td>
<td>alt1</td>
<td>3.10 (0.039)</td>
<td>alt1</td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>3.03 (0.059)</td>
<td>alt2</td>
</tr>
<tr>
<td>BF (no alt), new</td>
<td>alt1</td>
<td>3.10 (0.038)</td>
<td>alt1</td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>2.81 (0.70)</td>
<td>alt2</td>
</tr>
</tbody>
</table>

Table 12: Parameter values for fixed effects in mixed ordinal regression model of acceptability judgments in Experiment A.1

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>-0.8658</td>
<td>0.1515</td>
</tr>
<tr>
<td>Focus</td>
<td>-0.3615</td>
<td>0.1707</td>
</tr>
<tr>
<td>New:Focus</td>
<td>0.5338</td>
<td>0.2539</td>
</tr>
</tbody>
</table>

The same between-subjects manipulation of target identity was used as in Experiment A.1 to investigate the effect of the specific lexical material making up the target and the alternative expressions.

In the model including the between-subjects manipulation of target identity, the main effect of target identity did not reach significance ($z = 0.65$). However, this model revealed a significant three-way interaction between target identity, presence of an alternative and newness ($z = -2.55, p < 0.05$). Again, this indicates that acceptability judgments for items with alt2 as
Figure 4: Experiment A.1: mean rating in each condition. Error bars represent the 95% confidence interval.

the target were only significantly lower than items with alt1 as the target in the (NF) no alt, new condition ($z = -2.989, p < 0.05$ after Bonferroni correction for multiple comparisons).

1.3 Experiment A.3

Speaker A: This company often makes bad decisions, but...

a. Did they hire a lawyer last fall, or an accountant? (NF) alt, given
b. Did they hire a lawyer last fall? (NF) no alt, given
c. Did they hire an accountant last fall? (NF) alt, new
d. What did they announce last time? (NF) no alt, new

Speaker B: I think they announced it was a lawyer that they hired, but I’m not sure.

The data analysis was again analogous to that of Experiment A.2, except that it did not include a between-subjects fixed effects for target identity.
2 Appendix II: Materials

Materials for Experiment A.1 and Experiment 1 are given below. These materials were then adapted to create materials for the other experiments.

(1) **Context:** Abbie is a very picky eater.
   a. Did she want chocolate cake for dessert, or apple pie?  NF (alt), given
   b. Did she want chocolate cake for dessert?  BF (no alt), given
   c. Did she want apple pie for dessert?  NF (alt), new
   d. Do you remember what she said?  BF (no alt), new

**Target:** I think Abbie said she wanted chocolate cake for dessert, but I’m not sure.

(2) **Context:** Ben is feeling very sick and we’re trying to figure out why.
   a. Did he eat pasta at the restaurant, or pizza?  NF (alt), given
   b. Did he eat pasta at the restaurant?  BF (no alt), given
   c. Did he eat pizza at the restaurant?  NF (alt), new
### Table 13: Experiment A.2: mean rating and standard error of the mean by condition and by target and contextual alternatives.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Target Identity</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NF) alt, given</td>
<td>alt1</td>
<td>3.30</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>3.34</td>
<td>(0.048)</td>
<td></td>
</tr>
<tr>
<td>(NF) no alt, given</td>
<td>alt1</td>
<td>2.77</td>
<td>(0.039)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>2.82</td>
<td>(0.053)</td>
<td></td>
</tr>
<tr>
<td>(NF) alt, new</td>
<td>alt1</td>
<td>3.19</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>3.26</td>
<td>(0.046)</td>
<td></td>
</tr>
<tr>
<td>(NF) no alt, new</td>
<td>alt1</td>
<td>3.04</td>
<td>(0.040)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>alt2</td>
<td>2.89</td>
<td>(0.060)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 14: Parameter values for fixed effects in mixed ordinal regression model of acceptability judgments in Experiment A.2.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newness</td>
<td>-0.3194</td>
<td>0.1460</td>
<td>-2.188</td>
</tr>
<tr>
<td>Alternative</td>
<td>-1.4143</td>
<td>0.1661</td>
<td>-8.516</td>
</tr>
<tr>
<td>Newness:Alternative</td>
<td>0.9407</td>
<td>0.2306</td>
<td>4.079</td>
</tr>
</tbody>
</table>

d. What do you remember about yesterday?  
**BF (no alt), new**

**Target:** I think I saw him eating pasta at the restaurant, but it could have been something else.

(3) **Context:** We need a few computers for the lab.

a. Did Charlie buy a desktop at the store, or a laptop?  
**NF (alt), given**

b. Did Charlie buy a desktop at the store?  
**BF (no alt), given**

c. Did Charlie buy a laptop at the store?  
**NF (alt), new**

d. What did Charlie tell you again?  
**BF (no alt), new**

**Target:** I think Charlie told me he bought a desktop at the store, although I could be wrong.
Figure 6: Experiment A.2: mean rating in each condition. Error bars represent the 95% confidence interval.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Rating</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NF) alt, given</td>
<td>3.38</td>
<td>0.03</td>
</tr>
<tr>
<td>(NF) no alt, given</td>
<td>3.05</td>
<td>0.03</td>
</tr>
<tr>
<td>(NF) alt, new</td>
<td>3.27</td>
<td>0.03</td>
</tr>
<tr>
<td>(NF) no alt, new</td>
<td>2.86</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 15: Experiment A.3: mean rating and standard error of the mean by condition.

(4) **Context:** Dave had to get rid of a lot of his stuff.

a. Did he sell his washing machine when he moved out, or his dryer?   NF (alt), given
b. Did he sell his washing machine when he moved out?               BF (no alt), given
c. Did he sell his dryer when he moved out?                         NF (alt), new
d. What did he say about it?                                       BF (no alt), new

**Target:** I believe he said he sold his washing machine when he moved out, but he didn’t tell his roommates.
Figure 7: Experiment A.2: mean rating by alternative in each condition. Error bars represent the 95% confidence interval.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newness</td>
<td>-0.3790</td>
<td>0.1485</td>
<td>-2.552</td>
</tr>
<tr>
<td>Alternative</td>
<td>-0.9406</td>
<td>0.1503</td>
<td>-6.257</td>
</tr>
<tr>
<td>Newness:Alternative</td>
<td>-0.2012</td>
<td>0.2471</td>
<td>-0.815</td>
</tr>
</tbody>
</table>

Table 16: Parameter values for fixed effects in mixed ordinal regression model of acceptability judgments in Experiment A.3

(5) **Context:** I wonder how Erik is doing these days.

a. Does he have regrets from his previous marriage, or fond memories?  NF (alt), given
b. Does he have regrets from his previous marriage?  BF (no alt), given
c. Does he have fond memories from his previous marriage?  NF (alt), new
d. What did he say the other day?  BF (no alt), new

**Target:** I believe he said he has regrets from his previous marriage, but I’m not sure.

(6) **Context:** I’m looking for someone who can drop this off at work.
Figure 8: Experiment A.3: mean rating in each condition. Error bars represent the 95% confidence interval.

a. Are you going to the store today, or to the office?  
   NF (alt), given

b. Are you going to the store today?  
   BF (no alt), given

c. Are you going to the office today?  
   NF (alt), new

d. What did you decide to do?  
   BF (no alt), new

**Target:** I decided that I am going to the store today, but I might change my mind.

(7) **Context:** I’m just trying to figure out the logistics for this weekend.

a. Are you dropping people off at the train station tomorrow afternoon, or at the bus stop?  
   NF (alt), given

b. Are you dropping people off at the train station tomorrow afternoon?  
   BF (no alt), given

c. Are you dropping people off at the bus stop tomorrow afternoon?  
   NF (alt), new

d. What do you think?  
   BF (no alt), new

**Target:** I think I will be dropping people off at the train station tomorrow afternoon, but I can pick
you up wherever.

(8) **Context:** I don’t know what I should get.

  a. Are you drinking beer tonight, or wine? NF (alt), given
  b. Are you drinking beer tonight? BF (no alt), given
  c. Are you drinking wine tonight? NF (alt), new
  d. What do you think? BF (no alt), new

**Target:** I think I will be drinking beer tonight, but I don’t know about the others.

(9) **Context:** I wonder how the reimbursement process works.

  a. Would it be better to pay with cash tomorrow, or with card? NF (alt), given
  b. Would it be better to pay with cash tomorrow? BF (no alt), given
  c. Would it be better to pay with card tomorrow? NF (alt), new
  d. What did Andrew say? BF (no alt), new

**Target:** Andrew said it would be better to pay with cash tomorrow, although it doesn’t really matter.

(10) **Context:** Do you remember,

  a. did Faye order rice with her meal, or fries? NF (alt), given
  b. did Faye order rice with her meal? BF (no alt), given
  c. did Faye order fries with her meal? NF (alt), new
  d. what did Faye say just now? BF (no alt), new

**Target:** I believe she said she ordered rice with her meal, but we should ask her when she’s back.

(11) **Context:** Greg offered to help me move my stuff next weekend, but

  a. does he drive a car these days, or a van? NF (alt), given
  b. does he drive a car these days? BF (no alt), given
  c. does he drive a van these days? NF (alt), new
d. what did he say exactly? BF (no alt), new

**Target:** I believe he said he drives a car these days, but I would give him a call.

(12) **Context:** I’m thinking of buying Hana a birthday present.

a. Has she been a fan of fantasy since her teenage years, or of science fiction? NF (alt), given
b. Has she been a fan of fantasy since her teenage years? BF (no alt), given
c. Has she been a fan of science fiction since her teenage years? NF (alt), new
d. What did she say again? BF (no alt), new

**Target:** I think she said she has been a fan of fantasy since her teenage years, but I’m not sure.

(13) **Context:** I’m not sure what to get at the supermarket.

a. Does Jonathan like vanilla as an ice cream flavor, or strawberry? NF (alt), given
b. Does Jonathan like vanilla as an ice cream flavor? BF (no alt), given
c. Does Jonathan like strawberry as an ice cream flavor? NF (alt), new
d. What did Jonathan say before he left? BF (no alt), new

**Target:** I remember that he said he likes vanilla as an ice cream flavor, although I could be wrong.

(14) **Context:** I might have left my stuff at Kate’s place after the event yesterday.

a. Did she find a jacket last night, or a sweater? NF (alt), given
b. Did she find a jacket last night? BF (no alt), given
c. Did she find a sweater last night? NF (alt), new
d. What did she say again? BF (no alt), new

**Target:** I think she said she found a jacket last night, but I would give her a call.
(15) **Context:** I’m trying to find out about the dietary restrictions of our guests.

a. Has Logan been allergic to peanuts ever since she was little, or to seafood? NF (alt), given
b. Has Logan been allergic to peanuts ever since she was little? BF (no alt), given
c. Has Logan been allergic to seafood ever since she was little? NF (alt), new
d. What did Logan say last time? BF (no alt), new

**Target:** I believe she said she has been allergic to peanuts ever since she was little, but I will double check.

(16) **Context:** We have to update your immunization record before we can proceed.

a. Were you vaccinated for tetanus recently, or for chicken pox? NF (alt), given
b. Were you vaccinated for tetanus recently? BF (no alt), given
c. Were you vaccinated for chicken pox recently? NF (alt), new
d. What did your doctor say? BF (no alt), new

**Target:** I think my doctor said I was vaccinated for tetanus recently, although I could be wrong.

(17) **Context:** I’m just wondering who made such a mess on this table.

a. Did Maria read a newspaper this morning, or a magazine? NF (alt), given
b. Did Maria read a newspaper this morning? BF (no alt), given
c. Did Maria read a magazine this morning? NF (alt), new
d. What did Maria say? BF (no alt), new

**Target:** I think she said she was reading a newspaper this morning, but I’m not sure.

(18) **Context:** I’m looking for some recommendations.
a. Does Tony like to listen to music while driving to work, or to a podcast?  
   NF (alt), given
b. Does Tony like to listen to music while driving to work?  
   BF (no alt), given
c. Does Tony like to listen to a podcast while driving to work?  
   NF (alt), new
d. What did Tony say again?  
   BF (no alt), new

**Target:** He said he usually likes to listen to music while driving to work, but he has horrible taste.

(19) **Context:** I’m trying to gauge his background knowledge.

a. Did he study biology in high school, or chemistry?  
   NF (alt), given
b. Did he study biology in high school?  
   BF (no alt), given
c. Did he study chemistry in high school?  
   NF (alt), new
d. What did he tell you?  
   BF (no alt), new

**Target:** I remember that he said he studied biology in high school, but you should ask him yourself.

(20) **Context:** Oliver really was an annoying kid.

a. Did he always make fun of his mother when he was younger, or of his sister?  
   NF (alt), given
b. Did he always make fun of his mother when he was younger?  
   BF (no alt), given
c. Did he always make fun of his sister when he was younger?  
   NF (alt), new
d. What did his dad say again?  
   BF (no alt), new

**Target:** I think his dad said he always made fun of his mother when he was younger, but it wasn’t too bad.
(21) **Context:** I wonder how your mom got the information.

a. Did she talk to a nurse at the hospital, or to a doctor?  
   NF (alt), given

b. Did she talk to a nurse at the hospital?  
   BF (no alt), given

c. Did she talk to a doctor at the hospital?  
   NF (alt), new

d. What did she tell you?  
   BF (no alt), new

**Target:** I believe she said she talked to a nurse at the hospital, but I might be mistaken.

(22) **Context:** I’m not sure what to bring tomorrow night.

a. Are you making a main dish for the dinner party, or a dessert?  
   NF (alt), given

b. Are you making a main dish for the dinner party?  
   BF (no alt), given

c. Are you making a dessert for the dinner party?  
   NF (alt), new

d. What did you decide?  
   BF (no alt), new

**Target:** I think I decided to make a main dish for the dinner party, but I’m not really a good cook.

(23) **Context:** What is your plan for tomorrow?

a. Is your dad coming over for lunch tomorrow, or for dinner?  
   NF (alt), given

b. Is your dad coming over for lunch tomorrow?  
   BF (no alt), given

c. Is your dad coming over for dinner tomorrow?  
   NF (alt), new

d. What did your dad say?  
   BF (no alt), new

**Target:** I think he said he is coming over for lunch tomorrow, but I will check.

(24) **Context:** I was thinking of buying some wool for Liz.

a. Is she knitting a scarf for her granddaughter, or socks?  
   NF (alt), given

b. Is she knitting a scarf for her granddaughter?  
   BF (no alt), given

c. Is she knitting socks for her granddaughter?  
   NF (alt), new

d. What did she say yesterday?  
   BF (no alt), new
Target: I think she said she is knitting a scarf for her granddaughter, but I will ask her again.

(25)  **Context:** This road has been closed for quite a while now.

a. Are they building a bridge here, or a tunnel?  
   NF (alt), given

b. Are they building a bridge here?  
   BF (no alt), given

c. Are they building a tunnel here?  
   NF (alt), new

d. What do you know about the situation?  
   BF (no alt), new

Target: I think they are building a bridge here, but they will be done very soon.

(26)  **Context:** I wonder if Rachel already knows about the recent divorce in her family.

a. Did she call her aunt last week, or her uncle?  
   NF (alt), given

b. Did she call her aunt last week?  
   BF (no alt), given

c. Did she call her uncle last week?  
   NF (alt), new

d. What did she tell you last night?  
   BF (no alt), new

Target: I think she said she called her aunt last week, but I don’t think she knows anything.

(27)  **Context:** I’m not sure what is appropriate in this case.

a. Are you giving them money for their wedding, or a giftcard?  
   NF (alt), given

b. Are you giving them money for their wedding?  
   BF (no alt), given

c. Are you giving them a giftcard for their wedding?  
   NF (alt), new

d. What do you think?  
   BF (no alt), new

Target: I think I am giving them money for their wedding, but I might change my mind.

(28)  **Context:** There was an accident on the highway.

a. Does Stephanie take the bus to school every day, or the train?  
   NF (alt), given

b. Does Stephanie take the bus to school every day?  
   BF (no alt), given
Target: Her mom said Stephanie takes the bus to school every day, but I’m not sure.

(29) Context: I haven’t heard anything yet.

a. Did Dan receive a letter last month, or an email? NF (alt), given
b. Did Dan receive a letter last month? BF (no alt), given
c. Did Dan receive an email last month? NF (alt), new
d. What did Dan tell you? BF (no alt), new

Target: He told me he received a letter last month, but you should just give them a call.

(30) Context: I’m not sure when we should have our new furniture delivered.

a. Did you paint the walls this week, or the ceiling? NF (alt), given
b. Did you paint the walls this week? BF (no alt), given
c. Did you paint the ceiling this week? NF (alt), new
d. What did you decide? BF (no alt), new

Target: I decided to paint the walls this week, and I hope to be done with the first floor next week.

(31) Context: I must be going deaf!

a. Did you hear the door bell just now, or the microwave? NF (alt), given
b. Did you hear the door bell just now? BF (no alt), given
c. Did you hear the microwave just now? NF (alt), new
d. What did you say? BF (no alt), new

Target: I said I heard the door bell just now, but I might be wrong.

(32) Context: I’m updating the roster.

a. Did Tom choose to write a paper for this class, or to take the exam? NF (alt), given
b. Did Tom choose to write a paper for this class? BF (no alt), given

c. Did Tom choose to take the exam for this class? NF (alt), new

d. What did Tom say? BF (no alt), new

**Target:** I think Tom said he chose to write a paper for this class, but he could change his mind.

(33) **Context:** We should find a place to stay for next weekend.

a. Is Caroline renting a house in the city, or an apartment? NF (alt), given
b. Is Caroline renting a house in the city? BF (no alt), given

c. Is Caroline renting an apartment in the city? NF (alt), new
d. Do you remember what Caroline said? BF (no alt), new

**Target:** I remember Caroline said she is renting a house in the city, but we should ask her again.

(34) **Context:** We’re almost done with the side dishes, but

a. did Vera cut up cucumbers for the salad, or tomatoes? NF (alt), given
b. did Vera cut up cucumbers for the salad? BF (no alt), given

c. did Vera cut up tomatoes for the salad? NF (alt), new
d. what did Vera say? BF (no alt), new

**Target:** I think Vera said she cut up cucumbers for the salad, although it doesn’t really matter.

(35) **Context:** Wendy is not allowed to watch everything.

a. Did she watch a sitcom yesterday, or a documentary? NF (alt), given
b. Did she watch a sitcom yesterday? BF (no alt), given

c. Did she watch a documentary yesterday? NF (alt), new
d. What did she tell you? BF (no alt), new

**Target:** I believe she said she watched a sitcom yesterday, but I’m not sure.

(36) **Context:** Something is different here!

a. Did Saul move the table to the other side of the room, or the sofa?
b. Did Saul move the table to the other side of the room? BF (no alt), given

c. Did Saul move the sofa to the other side of the room? NF (alt), new

d. What did Saul say? BF (no alt), new

Target: He said he moved the table to the other side of the room, although I’m not sure if I like it.

(37) Context: This company often makes the wrong decisions.

a. Did they hire a lawyer last fall, or an accountant? NF (alt), given
b. Did they hire a lawyer last fall? BF (no alt), given

c. Did they hire an accountant last fall? NF (alt), new
d. What did they announce this time? BF (no alt), new

Target: I think they announced that they hired a lawyer last fall, but I might be wrong.

(38) Context: What are you doing for the holidays?

a. Are you celebrating new year’s with family this year, or with friends? NF (alt), given
b. Are you celebrating new year’s with family this year? BF (no alt), given
c. Are you celebrating new year’s with friends this year? NF (alt), new
d. What did you decide? BF (no alt), new

Target: I decided I will celebrate new year’s with family this year, but I might change my mind.

(39) Context: I’m making the same recipe as Zara did last time.

a. Did she use basil for the sauce, or parsley? NF (alt), given
b. Did she use basil for the sauce? BF (no alt), given
c. Did she use parsley for the sauce? NF (alt), new
d. What did she say? BF (no alt), new

Target: She said she used basil for the sauce, although I could be wrong.
(40) **Context:** The police are trying to find out how the burglar got in.

a. Did Amanda close the door when it got cold, or the window? NF (alt), given
b. Did Amanda close the door when it got cold? BF (no alt), given
c. Did Amanda close the window when it got cold? NF (alt), new
d. What did Amanda tell them? BF (no alt), new

**Target:** I think she said she closed the door when it got cold, but she didn’t lock it.

(41) **Context:** I’m making Bill’s schedule right now.

a. Is he teaching on Tuesdays this quarter, or on Wednesdays? NF (alt), given
b. Is he teaching on Tuesdays this quarter? BF (no alt), given
c. Is he teaching on Wednesdays this quarter? NF (alt), new
d. What did he tell you? BF (no alt), new

**Target:** I believe he said he will be teaching on Tuesdays this quarter, but I’m not sure.

(42) **Context:** I don’t know what the weather will be like.

a. Should I wear shorts today, or jeans? NF (alt), given
b. Should I wear shorts today? BF (no alt), given
c. Should I wear jeans today? NF (alt), new
d. What do you think? BF (no alt), new

**Target:** I think you should wear shorts today, but you should decide for yourself.

(43) **Context:** I wonder if we have all the ingredients already.

a. Do you still need milk for this recipe, or eggs? NF (alt), given
b. Do you still need milk for this recipe? BF (no alt), given
c. Do you still need eggs for this recipe? NF (alt), new
d. What do you think? BF (no alt), new
Target: I think we still need milk for this recipe, but I will check the fridge.

(44) **Context:** I’m trying to decide if I should make a reservation.

a. Would you like to sit in the back during the show, or in the front?  
   NF (alt), given

b. Would you like to sit in the back during the show?  
   BF (no alt), given

c. Would you like to sit in the front during the show?  
   NF (alt), new

d. What do you think?  
   BF (no alt), new

Target: I think I would like to sit in the back during the show, but you don’t have to make a reservation.

(45) **Context:** What was going on?

a. Was Jack looking for his wallet in the car, or for his keys?  
   NF (alt), given

b. Was Jack looking for his wallet in the car?  
   BF (no alt), given

c. Was Jack looking for his keys in the car?  
   NF (alt), new

d. What did Jack tell you?  
   BF (no alt), new

Target: I think he said he was looking for his wallet in the car, but he didn’t find anything.

(46) **Context:** We’re trying to give away the leftovers.

a. Did Claire bring the roasted vegetables to the potluck, or the fruit salad?  
   NF (alt), given

b. Did Claire bring the roasted vegetables to the potluck?  
   BF (no alt), given

c. Did Claire bring the fruit salad to the potluck?  
   NF (alt), new

d. What did Claire say?  
   BF (no alt), new

Target: I believe she said she brought the roasted vegetables to the potluck, but we should ask her again.

(47) **Context:** Yesterday the jewelry store was held up.
a. Did the thief steal a bracelet from the store, or a necklace?  
   NF (alt), given
b. Did the thief steal a bracelet from the store?  
   BF (no alt), given
c. Did the thief steal a necklace from the store?  
   NF (alt), new
d. What did your hear about it?  
   BF (no alt), new

**Target:** I heard that they stole a bracelet from the store, and it wasn’t a very expensive one.

(48) **Context:** We already did a lot of chores today!

   a. Did Dean do the dishes this morning, or the laundry?  
      NF (alt), given
   b. Did Dean do the dishes this morning?  
      BF (no alt), given
c. Did Dean do the laundry this morning?  
      NF (alt), new
d. What did Dean tell you?  
      BF (no alt), new

**Target:** He said he did the dishes this morning, but I’m not sure.