AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2023, Vol. 49, No. 12, 1971–1988
https://doi.org/10.1037/xlm0001292

In Conversation, Answers Are Remembered Better Than the Questions Themselves

Eirini Zormpa[1], Antje S. Meyer[1, 2], and Laurel E. Brehm[1, 3]
[1] Max Planck Institute for Psycholinguistics
[2] Radboud University Nijmegen
[3] Department of Linguistics, University of California, Santa Barbara

Language is used in communicative contexts to identify and successfully transmit new information that should be later remembered. In three studies, we used question–answer pairs, a naturalistic device for focusing information, to examine how properties of conversations inform later item memory. In Experiment 1, participants viewed three pictures while listening to a recorded question–answer exchange between two people about the locations of two of the displayed pictures. In a memory recognition test conducted online a day later, participants recognized the names of pictures that served as answers more accurately than the names of pictures that appeared as questions. This suggests that this type of focus indeed boosts memory. In Experiment 2, participants listened to the same items embedded in declarative sentences. There was a reduced memory benefit for the second item, confirming the role of linguistic focus on later memory beyond a simple serial-position effect. In Experiment 3, two participants asked and answered the same questions about objects in a dialogue. Here, answers continued to receive a memory benefit, and this focus effect was accentuated by language production such that information-seekers remembered the answers to their questions better than information-givers remembered the questions they had been asked. Combined, these studies show how people's memory for conversation is modulated by the referential status of the items mentioned and by the speaker's roles of the conversation participants.

Keywords: focus, question–answer pairs, recognition memory, dialogue, production effect

Supplemental materials: https://doi.org/10.1037/xlm0001292.supp

We are exposed to immense amounts of language every day, and for better or worse, we do not retain all of it. What information we do remember depends on many factors, including how the information is presented and on our own role in the discourse. Research on information structure shows that when information is presented as important in a discourse, otherwise known as *focused*, it tends to be remembered better than when the same information is presented neutrally or in contrast to focused information (see Birch & Garnsey, 1995; Cutler & Fodor, 1979; Fraundorf et al., 2010; Johns et al., 2014; Sturt et al., 2004). In the current work, we investigate how question–answer pairs, a naturalistic manipulation of focus, affect memory for items mentioned in conversation. We establish how focus affects memory for overheard conversations (Experiments 1 and 2) and how those patterns are modulated by engaging in the conversation (Experiment 3). This sheds light on the ways conversation filters what we remember about the world.

Focus has been researched extensively in the past decades and has been associated with a number of definitions and manipulations. Here, we follow Levelt (1989) and take focus to refer to the most attended part of a discourse. Focus can be induced in many ways, including by manipulating properties of words (Halliday, 1967), syntax (Birch & Garnsey, 1995; Birch & Rayner, 1997), or semantic context (Cutler & Fodor, 1979). Focused information has been associated with advantages in online processing and later memory. For instance, focused information tends to be processed for longer and in more detail (Benatar & Clifton, 2014; Birch & Garnsey, 1995; Birch & Rayner, 1997; Osaka et al., 2002; Ward & Sturt, 2007) and is remembered better than information that is neutral or not focused (Birch & Garnsey, 1995; Cutler & Fodor, 1979; Fraundorf et al., 2010; Johns et al., 2014; Sturt et al., 2004, though cf. Almor & Eimas, 2008). These phenomena have been argued to stem from a common cause: focused items are encoded more deeply than non-focused items, leading to stronger representations in the discourse model (Foraker & McElree, 2007). Indeed, Sturt et al. (2004) found that participants were more accurate at detecting changes in a text when the element that was changed had previously been focused than when it had not. Moreover, Wang et al. (2011) found larger N400 components, interpreted to reflect depth of processing, for focused items than nonfocused items. Focus has also

This article was published Online First November 30, 2023.
Laurel E. Brehm https://orcid.org/0000-0003-0424-8735
The preregistration, materials, data, and analyses for Experiments 1 and 2 are available at https://osf.io/x45ad/ (Zormpa et al., 2023a), and the preregistration, materials, data, and analyses for Experiment 3 are available at https://osf.io/y7seu/ (Zormpa et al., 2023b).

The authors thank Caitlin Decuyper and Dennis Joosen for their help with stimuli creation, Sanne van Eck and Christiane Ruhrmann for an assistance with participant testing, Christiane Ruhrmann for compiling and plotting the Experiment 3 eye-tracking data, Maarten van den Heuvel for programming and other technical support, and Birgit Knudsen for counting the occurrences of questions and clefts in the Ghent Eye-Tracking Corpus.

Correspondence concerning this article should be addressed to Eirini Zormpa, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, the Netherlands. Email: ezormpa@turing.ac.uk

been associated with visual attention: participants spend more time reading items that are focused than items that are not focused (Benatar & Clifton, 2014; Birch & Rayner, 1997; Lowder & Gordon, 2015, though cf. Birch & Rayner, 2010). Together, these findings suggest that focus causes people to process information more attentively and encode it more deeply, leading to stronger representations and more consistent knowledge retention.

However, we note that these manipulations of focus might not obtain in conversation. The most common sentence-level manipulations of focus are clefts and pseudo-clefts (e.g., Almor & Eimas, 2008; Birch et al., 2000; Birch & Garnsey, 1995; Birch & Rayner, 1997, 2010; Foraker & McElree, 2007; Järvikivi et al., 2014; Lowder & Gordon, 2015; Morris & Folk, 1998; Sanford et al., 2009), which are exceedingly rare structures and appear in $< 0.1\%$ of English sentences (Roland et al., 2007). In these structures, syntax guides attention to one element of the sentence, for example,"It is the goat that should move next to the painting" or "What should move next to the painting is the goat."

A more naturalistic way of inducing focus is the manipulation of the semantic context through questions. This also has been shown to elicit memory benefits in comprehension studies. Indirect question/answer pairs like "Everyone wanted to know which item should move. It turns out the goat should move next to the painting" improve memory for the focused item (Benatar & Clifton, 2014; Cutler & Fodor, 1979; Sauermann et al., 2013; Sturt et al., 2004; Wang et al., 2011; Ward & Sturt, 2007; Yang et al., 2017). Direct questions also elicit focus effectively (Chomsky, 1971), are very frequent in conversation (Graesser et al., 1994), and affect memory. Cutler and Fodor (1979) used auditory questions that put either the subject or the object of the answer in focus. In a four-alternative forced-choice sentence completion task, participants made fewer errors when the target response had been a focused item than a nonfocused item. Yang et al. (2017) used questions that varied whether the focus was placed on a word or not; at the end of a three-sentence-long narrative, that word acted as the target in a probe recognition task. Responses to those targets were faster after focusing questions than after nonfocusing questions.

However, these previous studies all tested how a preceding question affects memory for different parts of the answer relative to each other. To understand how these properties impact communication, it is also critical to know how answers are remembered relative to questions. This is especially important when considering both sides of a conversation because questions and answers are typically uttered by different people. Understanding more about how questions and answers are represented, therefore, gives us insight into the discourse models that different interlocutors build for conversation and how this translates to later memory. This is the goal of the current studies.

Furthermore, an important feature of conversation is that it typically involves participants taking turns speaking and listening. There is a known benefit of memory for speaking, compared to listening, and it is plausible that the linguistic focus may moderate this speaker benefit. Broadly speaking, the pattern is that speakers remember what they said better than listeners remember what they heard. This speaker benefit holds across a variety of stimuli: it has been explored for individual words and pictures (Brown et al., 1995; McKinley et al., 2017; Yoon et al., 2016), as well as for sentences (Jarvella & Collas, 1974; Miller, 1996) and the cues used to generate sentences (Fischer et al., 2015). The speaker benefit holds for recall tasks (Miller, 1996) but has been tested more often for recognition memory tasks (Fischer et al., 2015; Jarvella & Collas, 1974; McKinley et al., 2017; Yoon et al., 2016). It is long-lasting and can even be found 1 week after study (Brown et al., 1995).

The superior memory of speakers compared to listeners has been attributed to two effects associated with language production. These are the generation and production effects (Bertsch et al., 2007; Dew & Mulligan, 2008; MacLeod et al., 2010; Ozubko et al., 2014; Slamecka & Graf, 1978). The generation effect is the finding that coming up with a word provides a memory benefit relative to reading or hearing a word (Bertsch et al., 2007; Dew & Mulligan, 2008; Slamecka & Graf, 1978). The generation effect has been attributed to increased item-specific processing (Hunt & McDaniel, 1993) and to the relative distinctiveness of the resulting memory trace (Gardiner & Hampton, 1988). The production effect refers to the finding that saying words aloud improves memory relative to saying them silently, that is, in inner speech (MacLeod et al., 2010). This effect has also been attributed to distinctiveness: speaking provides additional distinctive sensory information (Ozubko et al., 2014). Both effects are likely to be in play in typical conversational circumstances because speaking typically involves both generating and producing utterances, which is why we collapse both effects under the broader term "speaker benefit effect."

Existing work shows that the speaker benefit effect diminishes or even reverses in more conversational contexts (Hjelmquist, 1984; Knutsen & Le Bigot, 2014; Stafford & Daly, 1984, though see Miller, 1996). In one study, Knutsen and Le Bigot (2014) asked participants to come up with a route that crossed certain points marked on a map. Participants had 20 min to complete the task, after which they were instructed to write down as much of the conversation as they could recall. The authors reported more reuse of self- compared to other-introduced referents (e.g., landmarks and street names) during the conversation, but no memory advantage for self- as opposed to other-introduced referents in the memory task. Hjelmquist (1984) gave participant pairs a topic (e.g., recent political events) and let them talk for 7 min. When participants were presented with sentences from this conversation 4 days later they were equally good at recognizing their own and their interlocutors' sentences correctly. Similarly, Stafford and Daly (1984) had participant pairs get to know each other and then write down as much as they could remember about the conversation. Here, participants recalled more of the information provided by their interlocutors in the earlier conversation than the information they had provided about themselves, reversing the typical speaker memory advantage. These discrepant findings demonstrate the need for carefully controlled studies that directly contrast the size of the speaker benefit effect in monologue and dialogue contexts.

A variable that might explain differences among studies of the speaker benefit effect is the speaker role associated with these different situations. This returns to the notion of focus: some speaker roles also serve to highlight information as important. In most studies investigating the generation or production effects (Bertsch et al., 2007; Dew & Mulligan, 2008; MacLeod et al., 2010; Ozubko et al., 2014; Slamecka & Graf, 1978), participants speak for the sake of performing an experimental task alone: there is no communicative intent involved in their speech. In other, more conversational studies, the participants' intention is to communicate information, like the order of pictures in a grid, or to give instructions (McKinley et al., 2017; Yoon et al., 2016), which might place emphasis on the speaker-produced information. Finally, in some more naturalistic studies, participants' intention is to get information from their interlocutors (Stafford & Daly, 1984), which might place emphasis on the other-produced information. This means that the focus may impact whether a speaker benefit is observed at all. The current studies have the secondary goal of showing whether the generation and production effects appear in memory for conversation and whether they are impacted by linguistic focus.

Current Study

In order to establish the role of linguistic focus in conversations, we began by testing the effect of questions compared to answers on memory for passive comprehension. This was the goal of Experiment 1, in which participants heard question–answer pairs uttered by two different speakers like "What should move next to the painting?" "The goat" and saw pictures of a goat, a painting, and an unrelated item (a doll). We predicted that the focused items (answers) would be remembered better than neutral items (questions). That is, we expected "goat" to be remembered better than "painting" when embedded in a question–answer pair.

In Experiment 2, we used the same paradigm but presented the study materials in simple declarative sentences like "There is a painting." "There is a goat." The goal of the study was to isolate the role of linguistic focus from any serial position effect, such as effects of recency or primacy. Under the assumption that focus was an important driver of later memory, we predicted that the memory benefit of the second item ("goat") would, therefore, be smaller than what we observed in Experiment 1.

In Experiment 3, we established how these findings were modulated by individuals' conversational roles. That is, we explored how the speaker benefit effect interacted with linguistic focus in conversations, disclosing how speaker role impacts memory for conversation. We predicted that focus would attenuate the speaker benefit effect, so that answers should be remembered well by both speakers and listeners. Experiment 3 also investigated the link between focus and visual attention. Earlier work attributes the beneficial effects of focus in comprehension (Benatar & Clifton, 2014; Foraker & McElree, 2007) and production (Ganushchak et al., 2014) to increased attention or processing time (though see Birch & Garnsey, 1995; Ward & Sturt, 2007). Visual attention to referents presented alongside sentences has been used to index the mental processes behind speaking (Ganushchak et al., 2014; Griffin & Bock, 2000) and listening (Altmann & Kamide, 1999; Cooper, 1974); see Huettig et al. (2011) for a review. It may be the case that more attention is necessary for speaking than for listening, given, for example, the high fixation rates to mentioned items when speaking by Griffin and Bock (2000) compared to lower rates of fixations to mentioned objects in passive listening by Cooper (1974; see also Sjerps & Meyer, 2015). Plausibly then, visual attention might support both the focus effect and the speaker benefit effect. We tested this by examining how visual attention, assessed through eye movement recording, moderated the relationship between focus and memory. This sheds light on the underlying mechanisms of each effect.

Experiment 1

In this experiment, we established the impact of focus, manipulated via question–answer pairs, on later memory for overheard conversations.

Method

Participants

Forty-eight native Dutch speakers (38 female) aged 18–30 ($M = 23$) were recruited from the Max Planck Institute for Psycholinguistics participant database. They received 8 euros for their participation. None disclosed any speech and language problems and all had normal or corrected-to-normal vision.

We selected a sample size of 48 participants by running a power analysis in which we simulated data with effect sizes ranging from 3% to 6% memory improvement. With 128 target items (384 in total, which is as many as we could find) 42 participants would give us 80% power to detect condition-level differences of 4% or greater. We tested 48 participants to have a balanced number of participants in each list. Ethical approval to conduct this study was given by the Ethics Board of the Social Sciences Faculty of the Radboud University.

Materials

All materials for this study can be found at https://osf.io/x45ad/.

Pictures. In the first phase of the experiment, 384 color photographs were used as stimuli (triplets of images per trial). Most ($N = 322$) were sourced from the BOSS picture database (Brodeur et al., 2010, 2014), but 62 came from other stimulus sets (Brady et al., 2008, 2013; Moreno-Martínez & Montoro, 2012), or Wikimedia Commons. A full list of the stimuli and their sources can be found in the online supplemental materials.

All pictures were normed for name agreement, familiarity, visual complexity (measured in JPEG size, see Machado et al., 2015), log $_{10}$ frequency, and length (measured in letters). This was done in stages. First, 387 pictures were normed for name agreement by 15 participants recruited from the Max Planck Institute for Psycholinguistics participant database in an online study. Pictures with $< 83\%$ name agreement were replaced with pictures from the BOSS database that were previously normed in Dutch (Decuyper et al., 2021). Familiarity norms for all pictures from the BOSS set were drawn from Brodeur et al. (2010, 2014), and the remaining pictures were normed for familiarity by eight native Dutch speakers employed at the Max Planck Institute for Psycholinguistics in a second online study. Estimates of log $_{10}$ frequency for all items were taken from the SUBTLEX-NL corpus (Keuleers et al., 2010). These measures were used to split pictures in two balanced sets A and B using Match (van Casteren & Davis, 2007): name agreement ($M_A = .93$, $M_B = .93$), familiarity ($M_A = 4.27$, $M_B = 4.26$), visual complexity ($M_A = 48, 660$, $M_B = 48, 310$), log $_{10}$ word frequency ($M_A = 2.21$, $M_B = 2.26$), and length ($M_A = 6.72$, $M_B = 6.83$). These sets were counterbalanced across lists such that in two lists, set A was used as targets and set B as foils, and in two lists, set A was used as foils and set B as targets.

Within sets A and B, three further subgroups were created using Match resulting in subsets A1, A2, A3, and B1, B2, and B3. These were used to assign pictures to the question, answer, and unmentioned conditions used in the study phase of the experiment. Subsets A3 and B3 were always used as unmentioned items and the question and answer conditions were assigned to subsets A1 and A2 or B1 and B2 across four counterbalanced lists. The three subsets were combined into trials pseudo-randomly with one item from each such that none of the three pictures were semantically related or started with the same phoneme.

Pictures were presented in 300 × 300 pixel resolution against a white (RGB: 255, 255, 255) background.

Study Phase. In each trial in the study phase, participants saw three pictures and heard a conversation snippet between two native Dutch speakers (one female and one male). This is depicted in Figure 1. Speakers were recorded using Shure SM10A microphones while participating in a version of the experiment in which two participants asked and answered questions about the position of objects on the screen. The recordings were then edited to remove static and normalized in volume using Audacity. Silences at the end of the recordings were removed, and silences at the beginning of recordings that were relatively long or short compared to the others were shortened or lengthened accordingly such that trials began with 1,132 ms silence on average (*SD* 255).

Test Phase. In the test phase, participants saw the most common Dutch name for each of the 384 pictures. This is depicted in Figure 1. These were presented one at a time centrally on a white background (RGB: 255,255,255) in Calibri font, size 45.

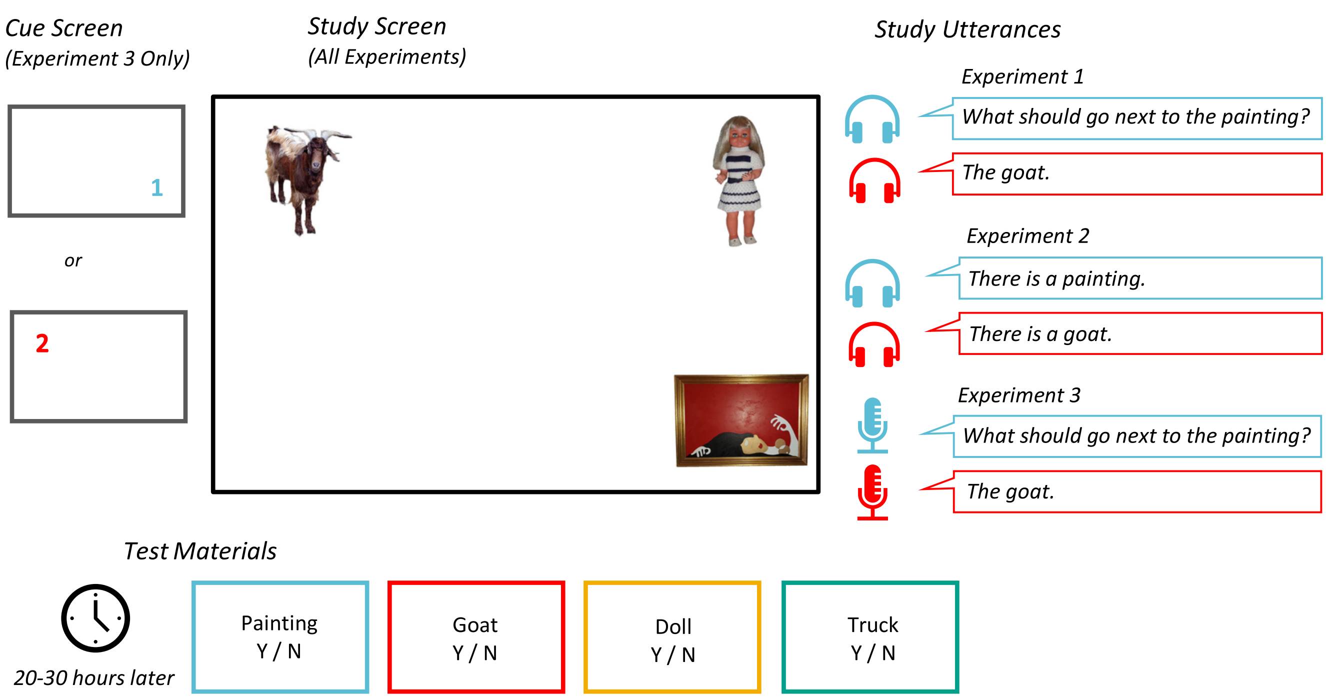
Procedure

In the study phase, participants were tested one or two at a time in separate soundproof booths in a session that lasted ~ 25 min. First, they completed four practice trials for which they received feedback and were encouraged to ask questions. Participants then completed 72 experimental trials where no feedback was provided. All trials started with a fixation cross displayed in the middle of the screen for 500 ms, followed by a blank screen that appeared for another 500 ms. Then, participants saw the three pictures, each occupying one of the four corners of the screen with one quadrant left blank. The position of the each item role (question, answer, unmentioned, and empty) was counterbalanced within lists and all combinations of role and location were used four times per list. The trials began with a silent period (duration 565–2,448 ms, $M = 1,132$ ms), followed by the conversation snippet (duration 2,460–4,052 ms, $M = 2,981$ ms). Participants then pressed on the space bar to move to the next trial.

Catch trials were also included to encourage participants to be attentive throughout the experiment. Two out of the four practice trials and eight out of the 72 experimental trials were catch trials. On these catch trials, participants were given a comprehension question after pressing the space bar to end the trial. The question queried the location of one of the item roles (question, answer, unmentioned, and empty). Participants had to answer the question by selecting one of the four corners using the keyboard. On the practice catch trials, feedback was provided, and on the experimental catch trials, no feedback was provided. Across all experimental catch trials, each role and each location was queried twice.

The second phase of the experiment was conducted online the following day. Participants were sent a link to an online yes/no recognition memory test and were given 8 hr to complete it. The names of all pictures that were shown the day before were presented, mixed with an equal number of foils. Participants were instructed to respond with "Yes" to the names of all pictures seen the previous

Figure 1
Schematic Diagram of Materials Used in Experiments 1, 2, and 3



Note. All participants viewed the same scenes paired with various utterances. Experiments 1 and 2 used prerecorded utterances (indicated by headphones) and Experiment 3 used a pair of participants producing the same utterances, (indicated by microphones), as cued by numbers presented on an earlier cue screen. All participants were tested on the same materials later: these were words that were either in the question (e.g. painting, blue outline), or answer (e.g. goat; red outline), named the unmentioned object (e.g. doll; yellow outline) or named foils that had not been shown in the study phase (e.g. truck; green outline). Headphone and microphone icons come from the Noun Project: https://thenounproject.com/. Photos come from the BOSS database (Brodeur et al., 2010, 2014). Y = yes; N = no. See the online article for the color version of this figure day, including those that were unmentioned. There was no time limit for the second session, but it usually lasted 10–15 min.

Analysis

Preregistered exclusion criteria included failure to complete the second phase of the experiment or below-chance performance in either phase of the experiment (under 25% of the catch trials in the first phase or under 50% in the memory task for the items in the questions and answers). No participants were excluded by these criteria.

Analyses were run using the lme4 package (Version 1.1-26; Bates et al., 2015) in R (Version 4.0.3; R Core Team, 2019) with the optimizer BOBYQA (Powell, 2009). The dependent variable were the log-odds of yes responses in the memory task. The random effects structure included random intercepts for items and participants, as well as any random slopes licensed for each random intercept. The exact structure was determined in a data-driven way starting from the maximal model and eliminating the slopes that explained the least variance if the model did not converge, or slopes that were correlated at a level of .95 or above with the random intercept. We report the final random effects structure in each model table.

Two preregistered analyses were run. The first aimed to ensure that participants were able to discriminate old from new items and used the target versus foil contrast as the only fixed effect. This was sum-to-zero contrast coded with targets coded as 0.5 and foils as −0.5. Next, the primary analysis tested the hypothesis that answers to questions are remembered better than questions. In this analysis, the answer versus question fixed effect was also sum-to-zero contrast coded with answers coded as 0.5 and questions as −0.5. The Open Science Framework preregistration can be found at: https://osf.io/x45ad/registrations.

Data and analyses can be found at: https://osf.io/x45ad/

Results

Accuracy in the catch trials at study was high ($M = 91\%$, $SD = 14\%$), meaning that participants were paying attention during the comprehension task. However, accuracy in the memory task was lower ($M = 58\%$, $SD = 49\%$). Performance for each item role in the memory task can be seen in Figure 2 (top panel). Participants were generally conservative: they were very good at correctly rejecting new items (19% false alarm rate), but they also falsely rejected many of the old items. Despite the relatively low accuracy, the pattern of responses to each condition follows the predicted pattern: recognition of old items was best for the answers and worst for the unmentioned items (Answers: $M = 40\%$, $SD = 49\%$; Questions: $M = 34\%$, $SD = 48\%$; Unmentioned: $M = 30\%$, $SD = 46\%$).

The first analysis examined whether participants could successfully distinguish old items (targets) from new ones (foils). The full logistic regression model for the analysis comparing targets and foils can be seen in Table 1. The random structure included by-participant and by-item intercepts and slopes for the target versus foil contrast (maximal model). The negative intercept shows that participants had a large *No* bias in this experiment. The positive estimate for the target versus foil comparison shows that participants responded positively to targets more often than to foils, that is, they were able to reliably distinguish old and new items at test.

The second analysis used only the trials in the memory task in which the items were verbally mentioned, as either the question or the answer in the study task. The full logistic regression model for this analysis is displayed in Table 2. The random structure included by-participant and by-item intercepts and by-item random slopes for the answer versus question contrast. Again, the negative intercept shows participants had a bias towardresponding *No* overall. The positive estimate for the answer versus question contrast shows that participants responded with *Yes* to answers more often than to questions. In other words, participants were more accurate when recognizing items that appeared as answers than items that appeared as questions.

Discussion

In this experiment, we tested whether the answers to questions were remembered better than the questions themselves. We predicted that memory for answers would be stronger than for questions because the answers are in focus. In line with this prediction, accuracy in the memory task was higher when a word had been used in an answer compared to when it had appeared in a question.

However, one confound presents itself: Answers by definition appear after the questions that elicit them. This creates a serial position effect within each dialogue, which could plausibly boost memory for the more recent item compared to the earlier item within each pair (see, e.g., Monsell, 1978; Neath, 1993 for serial position effects in recognition memory). To investigate the role of serial position in the observed memory boost from Experiment 1, we tested the same materials but presented them in simple declarative sentences rather than question–answer pairs. If linguistic focus has an effect beyond serial position, we predicted that Experiment 2 should see a reduced memory benefit for the item mentioned in the second position compared to Experiment 1.

Experiment 2

This experiment serves as a control for Experiment 1, dissociating the roles of focus and serial position in memory for overheard conversations.

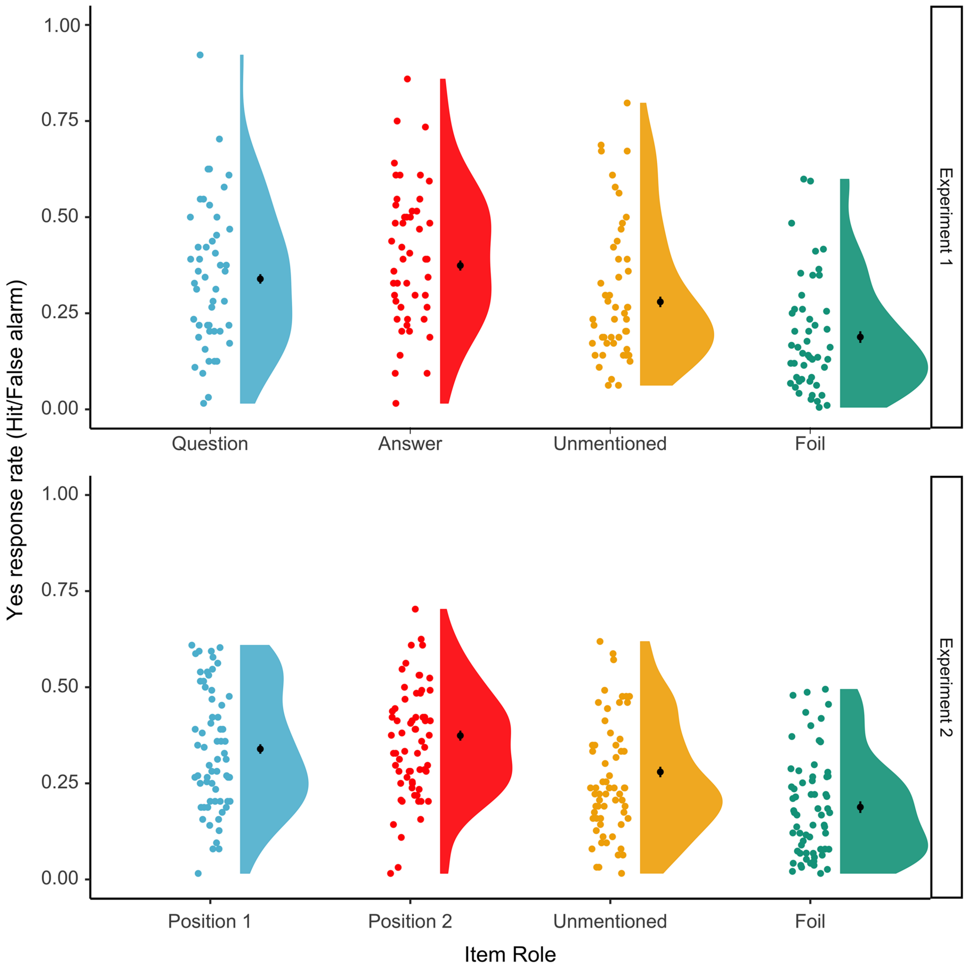
Method

Participants

Sixty-five native Dutch speakers (45 female) aged 18–33 ($M = 24$) were recruited from the Max Planck Institute for Psycholinguistics participant database; note that we extended the age limit slightly from our original preregistered criterion of age 30 because of recruitment difficulties. These participants received 8 euros for their participation. Three additional native Dutch speakers (three female) aged 18–30 were recruited as unpaid volunteers from another department of the Max Planck Institute for Psycholinguistics. No participants disclosed any speech or language problems and all had normal or corrected-to-normal vision. Of these 68 participants, four were excluded for completing their memory test more than 48 hr after the study session. Ethical approval to conduct this study was given by the Ethics Board of the Social Sciences Faculty of the Radboud University.

We selected a sample size of 64 participants based upon the power calculations in which we simulated data with varied participant and effect sizes and 128 target items. These showed that this sample size would be sufficient to detect a small simple effect of experiment (a

Figure 2
Yes *Responses to Each Item Role in the Memory Task for Experiment 1 (Top) and Experiment 2 (Bottom)*



Note. Colored points represent by-participant means and half-violins are distributions over participants. Black points are the grand means by condition, with 95% confidence intervals (CIs). See the online article for the color version of this figure.

2% difference in memory for the second object between Experiments 1 and 2) in a combined analysis of both experiments.

Materials

All materials for this study can be found at https://osf.io/x45ad/.

Pictures. The same pictures were used as in Experiment 1.

Study Phase. Participants saw the same displays as in Experiment 1 and again heard recorded utterances from the same two native Dutch speakers (one female and one male). As in Experiment 1, these utterances were described to participants as a conversation about objects on the screen. These were recorded and edited as described in Experiment 1. In addition, the length of the silent periods at the onset of each trial and between utterances was matched between experiments. See Figure 1 for a schematic diagram.

Test Phase. The test phase was identical to Experiment 1. See Figure 1 for a schematic diagram.

Procedure

The experimental procedure was nearly identical to Experiment 1. There were two differences. The most important difference was the content of the utterances in the study phase—simple declarative sentences like “There is a painting,” “There is a goat” rather than question–answer pairs. In addition, the catch trials queried the location of the item that was mentioned first, second, or unmentioned, or was empty, rather than querying the location of items in specific conversational roles.

Analysis

As in Experiment 1, no participants were excluded for the preregistered exclusion criterion of below-chance performance on catch trials. Our second preregistered exclusion criterion was for below-chance performance in the memory task. However, we noted that overall memory performance was fairly poor in this experiment: the average

Table 1
Mixed-Effects Logistic Regression Testing the Effect of Probe, That Is, Targets Versus Foils, in Experiment 1

Random effects	Variance	Correlation			
Item					
Intercept	0.12				
Target versus foil	0.37	−.48			
Participant					
Intercept	0.94				
Target versus foil	0.33	.57			
Fixed effects	**Estimate**	***SE***	**Wald *z***	***p***	**CI**
Intercept	−1.31	0.14	−9.12	<.001	[−1.59, −1.02]
Target versus foil	1.13	0.10	11.35	<.001	[0.94, 1.34]

Note. CI = confidence interval; *SE* = standard error.

memory performance was only 56% (range: 48%–61%). We therefore opted to include all participants in our analyses.

Two preregistered analyses and one exploratory analysis were run. The first compared the data from Experiments 1 and 2 using a generalized linear mixed model on the log-odds of yes responses in the memory task. In this analysis, item position was sum-to-zero contrast coded (first position: −0.5 and second position: 0.5) and experiment was dummy coded, with Experiment 1 as the baseline. This allows us to test for the simple effects of item position within Experiments 1 and 2. We made this choice because power simulations showed that this contrast coding was more suitable to observe the critical interaction of item mention with sufficient power at the chosen sample size of 65 participants. Random effects were determined as described in Experiment 1. We deemed that frequentist techniques were appropriate for this analysis (and the other analyses in the article) because the research question was focused on significance testing.

The second analysis examined only Experiment 2 data in a Bayesian logistic mixed effects regression with a weakly informative Cauchy prior (0, 2.5) on the effect of item position. This allowed us to estimate the mode of the posterior distribution in each condition, disclosing the likely effect size of first versus second mention of items. We deemed that Bayesian techniques were appropriate for this analysis because the research question was focused on directly comparing two effect sizes. These analyses were performed using the brms package, Version 2.16.1 (Bürkner, 2017). The final model used four chains with 8,000 iterations each and the first 4,000 iterations of each chain were discarded as a burn-in period; all parameters in this model attained an R-hat of 1.00 and no chain was divergent via visual inspection.

A third exploratory analysis, suggested by a reviewer, examined serial position effects in Experiments 1 and 2 in terms of trial order at study. While Experiment 1's results might be consistent with recency effects within each dialogue, primacy effects are more often observed in other literature using similar paradigms (Benjamin et al., 1998; Postman & Phillips, 1965). To examine whether the two experiments showed any reliable evidence for primacy or recency effects at study, we assessed whether adding trial order (centered) as linear and quadratic terms, either as main effects or in interaction with the other predictors, disclosed any reliable effects when added to the first preregistered analysis. The other predictors (item position and experiment) and random effects were as described for the first analysis.

The preregistration of this experiment appears at https://osf.io/mexwk/, while data and analyses can be found at: https://osf.io/x45ad/.

Results

Accuracy in the catch trials at study was again high ($M = 93\%$, $SD = 10\%$) and quite comparable to Experiment 1, showing that participants paid sufficient attention during the comprehension

Table 2
Mixed-Effects Logistic Regression Testing the Effects of Focus (Answers vs. Questions) in Experiment 1

Random effects	Variance	Correlation			
Item					
Intercept	0.09				
Answer versus question	0.07	.48			
Participant					
Intercept	0.87				
Fixed effects	**Estimate**	***SE***	**Wald *z***	***p***	**CI**
Intercept	−0.63	0.14	−4.53	<.001	[−0.91, −0.35]
Answer versus question	0.27	0.06	4.51	<.001	[0.15, 0.39]

Note. CI = confidence interval; *SE* = standard error.

task. While overall accuracy was slightly lower than in Experiment 1, the relative pattern of performance across conditions was similar (Position 2: $M = 36\%$, $SD = 48\%$; Position 1: $M = 34\%$, $SD = 47\%$; Unmentioned: $M = 26\%$, $SD = 44\%$). This is shown in Figure 2 (bottom panel). Participants were again generally conservative, which led them to successfully reject foils (again a 19% false alarm rate), and to incorrectly reject many of the items shown at study. Recognition of old items was again best for the items mentioned second and worst for the unmentioned items.

The first analysis, combining Experiments 1 and 2 together, appears in Table 3. This analysis included by-participant and by-item random intercepts and by-item random slopes for the first versus second mention contrast. Replicating what we found in Experiment 1, the negative intercept shows participants had a bias toward responding *No* in Experiment 1. Importantly, this analysis also showed a reliable interaction of item position and experiment: the second-position item was remembered better in Experiment 1 than in Experiment 2 (40% vs. 36%), while the first-position item was remembered equally well in both experiments (34%). This interaction was further supported by a simple effect of item mention in the baseline condition, Experiment 1. This reliable difference confirms that linguistic focus has a reliable effect on memory beyond what would be expected from recency alone. A visualization of Experiment 2 results appears in the bottom panel of Figure 2: note that the peak of the second-position item (red in online version) is shifted downward in Experiment 2 compared to Experiment 1.

The second analysis, examining effects in Experiment 2 alone, appears in Table 4 and included by-participant and by-item random intercepts and random slopes for the first versus second mention contrast. The 95% credible interval (CrI) for the effect of item position was relatively wide and included zero, consistent with the weak effect of this factor in the frequentist analysis, and the posterior estimate was 0.11. This suggests that most of the effect observed in Experiment 1 (with an estimate of 0.27 for the effect of questions vs. answers) was indeed due to linguistic focus.

A third exploratory analysis examined the effect of trial order at study as a measure of serial position effects. In this analysis, there were several significant effects involving the linear trial order effect (associated with increases or decreases in performance from beginning to end of experiment), and no effects involving the quadratic trial order effect (associated with an increasing, then decreasing, or decreasing, then increasing pattern). First, there was a significant interaction between item position and the linear trial order effect ($\beta = 17.81$, $SE = 4.79$, $z = 3.72$, $p < .001$). This showed that in Experiment 1 (the baseline condition), there was a decline in memory performance across trials for first-mentioned objects but not second-mentioned objects. Next, there was a significant interaction between experiment and the linear trial order effect ($\beta = 9.34$, $SE = 4.11$, $z = 2.27$, $p < .05$). This showed that in Experiment 2, there was a steeper decline in memory performance across trials than in Experiment 1. Finally, there was a significant interaction between item position, experiment, and the linear trial order effect ($\beta = -20.26$, $SE = 5.67$, $z = -3.57$, $p < .01$). This showed that in Experiment 2, the decline in memory performance across trials appeared for items in both positions. These results, therefore, suggest there is stronger evidence for the primacy than recency effects in the present data as a whole. Recency contributes relatively little to the memory benefit for the second item observed in Experiment 1, and memory performance decreases as trial order increases, especially without the protective effects of focus.

Discussion

Experiment 2 replaced the question–answer pairs used in Experiment 1 with pairs of simple declarative sentences. As confirmed by frequentist and Bayesian statistics, this reduced the memory advantage for the second item. These data allow us to rule out a simple item order effect as an explanation for Experiment 1's results: focus has an effect on memory that is not isolable to serial position within the conversational snippet.

Experiment 3

In the final experiment, we extended Experiment 1 to an interactive context where two speakers participated in a conversation rather than just observing it. Our primary aim was to establish the extent to which linguistic focus interacts with the speaker benefit in memory for conversations, shedding light on earlier conflicting evidence regarding what speakers and listeners remember from conversations. These preregistered analyses complement the questions asked in Experiment 1. We also explored the link between focus, visual attention, and memory. In a set of exploratory analyses, we examined patterns of object inspection in the study phase of the experiment and

Table 3
Mixed-Effects Logistic Regression Testing the Effect of Item Position, That Is, First Versus Second, by Experiment

Random effects	Variance	Correlation			
Item					
Intercept	0.15				
First versus second mention	0.02	.35			
Participant					
Intercept	0.67				
Fixed effects	**Estimate**	***SE***	**Wald *z***	***p***	**CI**
Experiment 1 performance (intercept)	−0.63	0.12	−5.09	<.001	[−0.88, −0.39]
First versus second mention (in Experiment 1)	0.28	0.06	4.77	<.001	[0.16, 0.39]
Experiment 2 performance	−0.11	0.16	−0.67	.50	[−0.42, 0.21]
First versus second mention (in Experiment 2 vs. Experiment 1)	−0.18	0.08	−2.40	<.005	[−0.33, −0.03]

Note. Note that Experiment 1 performance is mapped to the intercept term because of the contrast coding. CI = confidence interval; *SE* = standard error.

Table 4
Bayesian Mixed-Effects Logistic Regression Testing the Effect of Item Position, That Is, First Versus Second, in Experiment 2

Group-level effects	Estimate	*SE*	CrI	Bulk ESS	Tail ESS
Item					
Intercept	0.43	0.04	[0.36, 0.50]	7,816	11,217
First versus second mention	0.14	0.10	[0.01, 0.36]	4,313	7,155
Correlation	0.16	0.46	[−0.84, 0.93]	18,906	9,373
Participant					
Intercept	0.75	0.08	[0.61, 0.91]	4,808	9,108
First versus second mention	0.13	0.08	[0.01, 0.29]	5,792	7,102
Correlation	−0.45	0.40	[−0.98, 0.58]	17,643	10,206
Population-level effects	**Estimate**	***SE***	**CrI**	**Bulk ESS**	**Tail ESS**
Intercept	−0.74	0.10	[−0.94, −0.54]	3,224	5,537
First versus second mention	0.11	0.06	[0.00, 0.22]	21,733	12,051

Note. *SE* = standard error; CrI = 95% credible interval; ESS = effective sample size.

tested whether gaze durations at study were a reliable predictor of memory accuracy at test.

Method

Participants

A total of 110 participants were recruited in pairs. Of these, 103 were recruited from the Max Planck Institute for Psycholinguistics participant database. The first 32 participants were compensated 10 euros for their participation, and the remaining 71 were compensated 15 euros due to a change in participant payment policy associated with the COVID-19 pandemic. Seven additional participants were recruited as unpaid volunteers from other departments of the Max Planck Institute for Psycholinguistics. All participants (89 female) were 18–39 years old ($M = 23$) and were native Dutch speakers with no reported speech or language problems and with normal or corrected-to-normal vision. This extends the age limits reported in our preregistration; as in Experiment 2, we did this due to recruitment issues.

Of the 110 participants, 14 were excluded for the following reasons: two participants (in two different pairs) did not perform the task correctly and they and their partners were therefore excluded, two participants (one pair) had a computer crash, five participants (in five different pairs) completed the memory test more than 48 hr after the study phase, two participants (in different pairs) responded over 90% of the time with a single response (either yes or no), and one participant had less than chance accuracy.[1] This left a total of 96 participants contributing data to the experiment. This was our target sample size, determined by running a power analysis in which we simulated data using an effect size of ∼ 20% for the combined effect of production and generation and of 6% for the effect of focus, following Experiment 1. With those parameters, we would have 80% power to observe a significant interaction showing that information-seekers remember answers better than information-givers remember questions.

Ethical approval to conduct this study was given by the Ethics Board of the Social Sciences Faculty of the Radboud University.

Materials

Materials can be found at https://osf.io/y7seu/.

Pictures. The same pictures were used as in Experiments 1 and 2.

Study Phase. Participants were cued to produce question–answer pairs that were similar to those used in Experiment 1.

Test Phase. The test phase was identical to Experiments 1 and 2.

Procedure

The study phase was conducted in an experiment room at the Max Planck Institute for Psycholinguistics. Participants were seated side by side, each 55 cm in front of their own monitor (1,920 × 1,080 resolution) and separated by a cubicle divider. They were able to hear but not see each other. The experiment was controlled using Presentation (Version 18.3; Neurobehavioral Systems, Berkeley, CA, USA). Vocal responses from both participants were recorded using a directional head-mounted microphone. Eye data were collected using an EyeLink 1000 Plus (SR Research Ltd., Osgoode, Canada) eye-tracker.

At the beginning of the session, participants gave informed consent and then completed random-order nine-point calibration and validation routines. They then completed eight practice trials followed by 64 experimental trials. Participants received feedback during the practice trials, which were otherwise identical to the experimental trials. Participants had the option of a short break after 32 experimental trials.

The trial structure guided participants to produce utterances like those presented in Experiment 1. As shown in Figure 1, each trial started with an instruction screen, which consisted of a blue "1" for the information-seeker and a red "2" for the information-giver. These varied across trials, so that each participant was the information-seeker on 32 trials, and the information-giver on 32 trials. The position in which those numbers appeared signaled what items speakers should use in their utterances: each participant needed to use the name of the picture in their location. After 2,500 ms, the instruction screen was replaced by the experimental screen, consisting of three images and an empty quadrant. This screen was identical for both speakers. The information-seeker had to formulate a question about a new object

[1] The 90% single response criterion was determined after preregistration, as we noticed that some online participants in other studies in our group were not providing useful data but seemed to just be clicking through the experiment.

moving into the blank space in relation to the known object, eliciting utterances like, "What should move next to the painting?" The information-giver would then give the answer, which was their object ("The goat"). Finally, both participants pressed the space bar. The following trial started when the experimenter pressed a button on a button box. This was done to avoid the two systems going out of synchronization. Trial order was randomized with a unique random trial list presented to each pair of participants.

The test phase of the experiment was conducted online the following day, and followed the same procedure outlined in Experiments 1 and 2.

Analysis

The dependent variable in all analyses was memory performance ("Yes": 1; "No": 0). Memory performance was analyzed using logistic mixed-effects regression models run in R (Version 3.6.1; R Core Team, 2019) and implemented in the lme4 package (Version 1.1-21; Bates et al., 2015) using the optimizer *BOBYQA* (Powell, 2009). There were two confirmatory analyses and one exploratory analysis, which were preregistered on the Open Science Framework (https://osf.io/y7seu/registrations), as well as a second exploratory analysis added after data collection.

Predictors in the main confirmatory analysis were the participant's speaker role (information-seeker vs. information-giver) and item condition (self-produced, other-produced, and unmentioned). The speaker role predictor assessed the effect of the communicative role assigned to the participant in a particular trial (to gain or to give information), and the item predictor tested the effects of speaker benefit (self- vs. other-produced) and of item mention (mentioned vs. unmentioned). Examining the effect of speaker role, instead of the effect of question versus answer, makes the two factors fully independent because the simple effect of speaker role in the other-produced condition is equivalent to the question versus answer comparison made in Experiment 1. This analysis was run on the targets only, as foils did not belong to any of these conditions. A separate confirmatory analysis examined the effect of probe type (target vs. foil) to ensure that participants were able to distinguish between old and new items.

An additional exploratory analysis was run on all targets to assess the influence of visual attention, as measured by gaze duration, on memory. This analysis used all the predictors from the main analysis plus the continuous predictor gaze duration, which was operationalized as the total amount of looking time to a 400×400 pixel region surrounding the object photo during the time interval that it appeared on the screen (from the onset of the study screen to the offset of the trial). Data were excluded for two participants who could not be successfully calibrated.

Trials were excluded from all analyses when a picture was named incorrectly by either participant in the pair. In the cases when the incorrect name appeared elsewhere in the experiment, both instances of the word were excluded. This led to the exclusion of 607 trials (out of 6,144; 9.9% of the data). An additional 466 trials were excluded from the eye-tracking analyses because no fixations were registered to the object interest areas during the trial; we attribute these to fixation drift during the trial and to occasional participant inattentiveness.

Data and analyses can be found at: https://osf.io/y7seu/.

Results

Memory performance overall was 64% (*SD*: 48%). Consistent with Experiments 1 and 2, answers were remembered better than questions, which were in turn remembered better than unmentioned items. More specifically, words that appeared as answers were recognized correctly 59% (*SD*: 49%) of the time, words that appeared as questions 49% (*SD*: 50%) of the time, and words that appeared as unmentioned items 24% (*SD*: 40%) of the time. A visualization of the results can be seen in Figure 3.

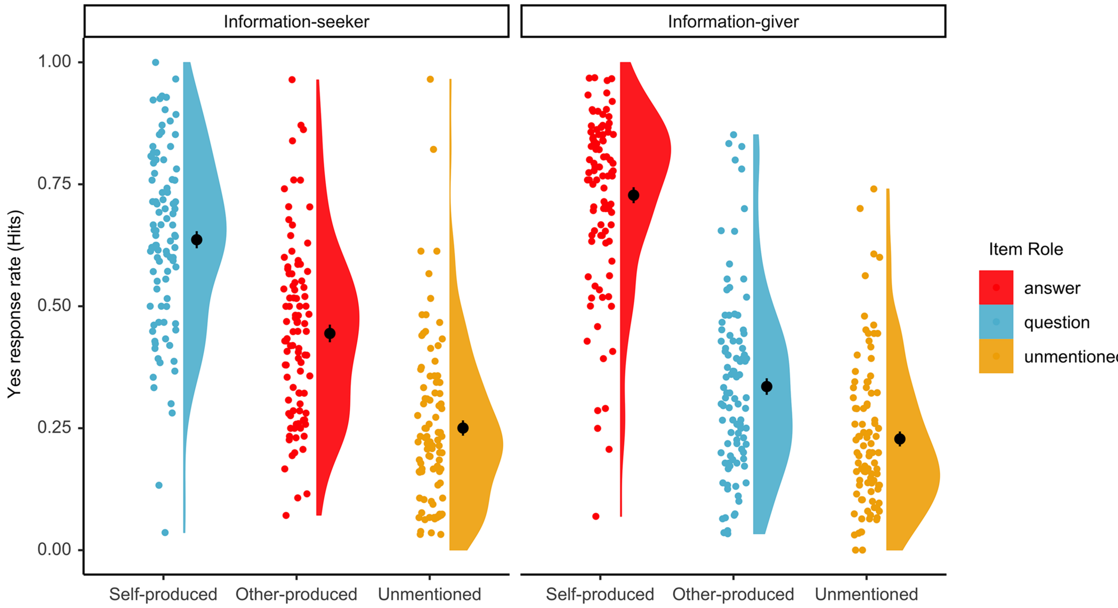
The first confirmatory analysis tested the effect of probe type (deviation contrast coded as targets =0.5 and foils =−0.5) to ensure that participants could distinguish between old and new items. The random structure included by-participant and by-item intercepts and by-participant and by-item slopes for the targets versus foils contrast. The results of this analysis can be seen in Table 5. The negative intercept suggests that participants had a negative response bias: as in the first two experiments, participants were overall more likely to respond "No" than "Yes." The positive estimate for the targets versus foils contrast suggests that participants were more likely to respond with "Yes" to targets than to foils: they were reliably able to distinguish between old and new items.

The second confirmatory analysis tested how speaker role and item condition affected memory for conversations. Speaker role was deviation contrast-coded (information-seeker =0.5, information-giver =−0.5). Item condition was Helmert contrast-coded and split into two contrasts: one testing the effect of having been mentioned (by self and by other =0.25, by no one =−0.5) and one testing production (by self =−0.5, by other =0.5). The random structure for this model included by-participant and by-item intercepts, random slopes by participant for speaker role and item condition, and random slopes by item for item condition. A visualization of the underlying data can be seen in Figure 3.

Broadly speaking, self-produced items were remembered best, and the effect of focus modulated this pattern so that answers were remembered better than questions when produced by another speaker. This is shown statistically in Table 6, which we now unpack. The negative intercept indicates a slight "No" response bias, which is consistent with the fact that the unmentioned items, included in this analysis, were associated with very low recognition rates. The nonsignificant effect of speaker role shows that there was no reliable difference between information-seekers and information-givers in how well the three objects were remembered overall ("Yes" rates of 44% for the information-seekers, and 43% for the information-givers). The positive estimate for the mention contrast shows that the mentioned items, regardless of who they were mentioned by, were remembered reliably better than the unmentioned items (the yellow distributions in each panel in color version of Figure 3). The negative estimate for the production contrast shows that participants were better at recognizing items they named themselves, regardless of whether they were questions or answers (the left-most distribution in each panel in Figure 3). This is evidence for the speaker benefit effect. The nonsignificant interaction between speaker role and mention shows that there was no reliable difference in how well unmentioned items were remembered by information-seekers as opposed to information-givers: the memory benefit of seeking information is restricted to items in focus. Finally, the key interaction between speaker role and production was significant. This shows that there is a difference between how well other-produced words are remembered by information-seekers and information-givers. Information-seekers (who asked the questions) remembered the answers better than information-givers (who gave the answers) remembered the questions. This provides an evidence for focus modulating the speaker benefit effect.

Figure 3

Hit Rates in the Experiment 3 Memory Task to Each Item Role Split by Speaker Role



Note. The information-seeker (in the left panel), asks questions. They remember the self-produced item in the question better than the answer. The information-giver (in the right panel) provides answers. They remember the self-produced item in the answer better than the question. That is, both speakers benefit from the generation effect and the production effect. However, the difference in hit rates for the self- and other-produced items is smaller in the case of the information-seeker, because the focus makes the answer more memorable. The violins are colored to represent the role of the item in the trial (as answer, question, or unmentioned). Each colored point represents a participant's mean hit rate for that condition. The black points represent the overall mean hit rate for that condition. The bars around the black point represent the normalized within-participant 95% confidence interval. See the online article for the color version of this figure.

A further exploratory analysis examined the relationship between visual attention at study and memory performance at test. We begin by describing the qualitative pattern of visual attention at study. Participants tended to first look at their own object (doing so on 96% of trials) and then look at the other-mentioned object once it became conversationally relevant (doing so on 85% of trials). The average first fixation to the participant's own object within the trial period occurred at 403 ms (410 ms for information-givers and 395 ms for information-seekers); however, note that the paradigm directed participants' attention to their own object before the measurement period started, meaning that these are likely to reflect the onset of a second distinct fixation to the same interest area. For information-givers, the average first fixation to the question interest area was at 2,232 ms, and for information-seekers, the average first fixation to the answer interest area was at 3,166 ms. Participants fixated the unmentioned object on 74% of trials, and the average first fixation time to this region was 1,757 ms for information-givers and 2,972 ms for information-seekers. Participants made a median

Table 5

Mixed-Effects Logistic Regression Testing the Effect of Probe (Targets vs. Foils) in Experiment 3

Random effects	Variance	Correlation			
Participant					
Intercept	0.43				
Targets versus foils	0.23	−.35			
Item					
Intercept	0.10				
Targets versus foils	0.25	−.26			
Fixed effects	**Estimate**	***SE***	**Wald *z***	***p***	**CI**
Intercept	−0.96	0.07	−13.63	<.001	[−1.10, −0.82]
Targets versus foils	1.37	0.06	22.23	<.001	[1.25, 1.49]

Note. CI = confidence interval; *SE* = standard error.

Table 6
Mixed-Effects Logistic Regression Testing the Effect of Speaker Role, That Is, Information-Seeker Versus Information-Giver and Item Condition

Random effects	Variance	Correlation			
Participant					
Intercept	0.57				
Speaker role	0.02	−.15			
Mention	0.55	.03	−.06		
Production	0.46	.02	.34	−.88	
Item					
Intercept	0.2				
Mention	0.48	.43			
Production	0.16	−.14	−.70		
Fixed effects	**Estimate**	***SE***	**Wald *z***	***p***	**CI**
Intercept	−0.31	0.08	−3.69	<.001	[−0.47, −0.14]
Speaker role	0.06	0.04	1.51	.13	[−0.02, 0.14]
Mention	2.09	0.10	20.41	<.001	[1.87, 2.29]
Production	−1.49	0.09	−15.69	<.001	[−1.65, −1.32]
Speaker role: mention	−0.17	0.11	−1.59	.11	[−0.38, 0.04]
Speaker role: production	1.05	0.09	11.74	<.001	[0.87, 1.22]

Note. The item condition predictor was split into mention (mentioned vs. unmentioned) and production (self-produced vs. other-produced). CI = confidence interval; *SE* = standard error.

of four fixations to each of the answer and the question interest areas, and a median of one fixation to the unmentioned object interest area.

We calculated gaze durations by summing the total duration of all fixations to an interest area on trials where at least one fixation was registered to any interest area. On average, participants spent 1,850 ms looking at the question object, 2,055 ms looking at the answer object, and 650 ms looking at the unmentioned object. Since trials were on average 5,054 ms long, this means that about 90% of the trial period was typically spent looking at the objects on the screen. Overall gaze durations also differed slightly by subject role: information-seekers tended to spend slightly less total time looking at the three relevant objects than the information-givers did (1,525 ms vs. 1,635 ms total).

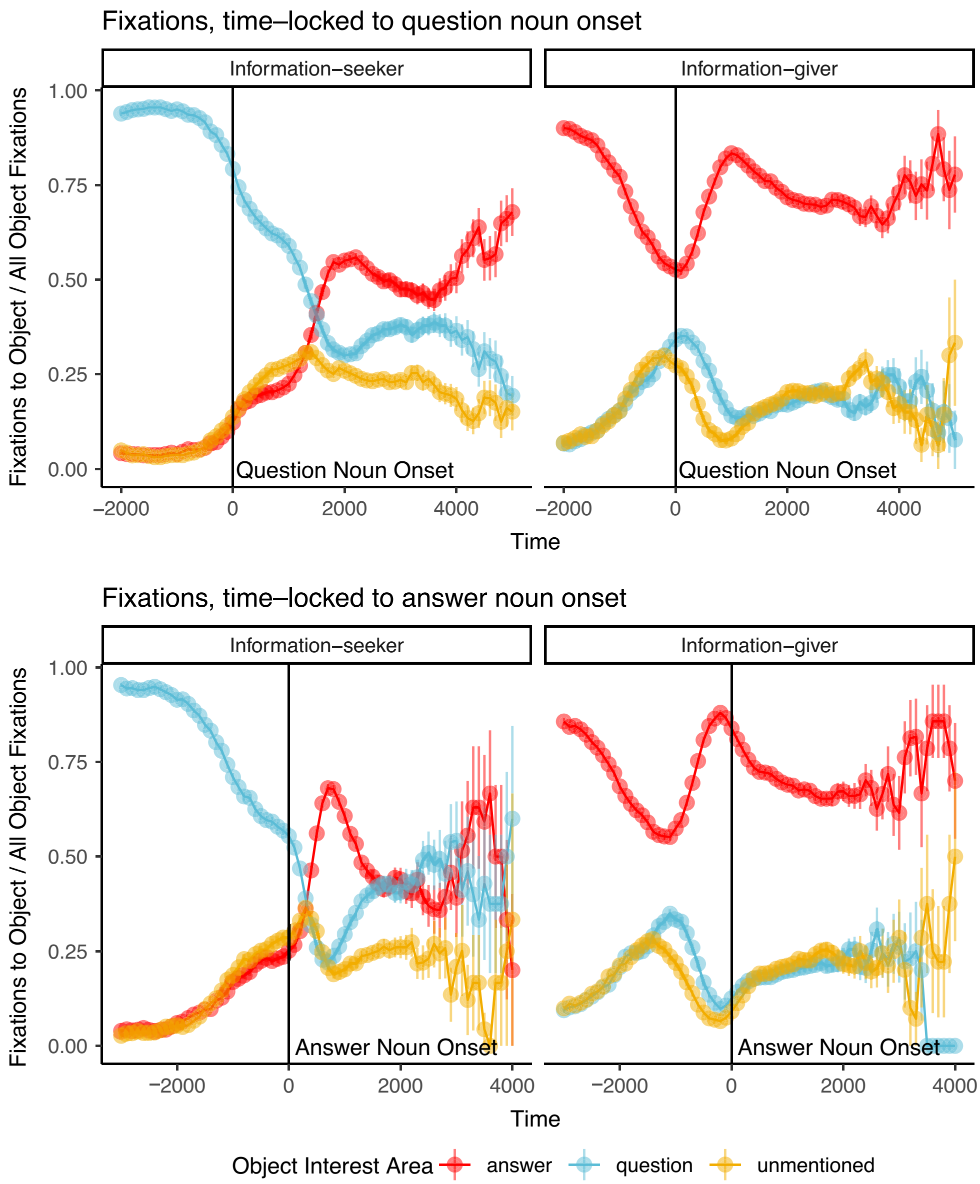
Fixations also followed an orderly relationship with the timing of speaking and listening. We quantified this by calculating the proportion of time within successive 100 ms time windows that participants spent fixating each interest area. We identified the peak of fixations to objects by finding the 100 ms time window where there was the highest proportion of fixations to each object for each subject role, as portrayed in Figure 4. For the information-seeker, the peak of fixations to the question object was 1,400 ms before production of the question noun, while for the information-giver, it was 100 ms after production of the question noun. For the information-giver, the peak of fixations to the answer object was 300 ms before production of the answer noun, while for the information-seeker, it was 600 ms after the production of the answer noun. These patterns are roughly comparable to earlier work using visual-world eye-tracking in language production and comprehension: attention is fixated to objects in preparation for speaking (Griffin & Bock, 2000) and in response to listening (Spivey et al., 2002), and it takes more time to prepare for speaking than to respond in listening (Sjerps & Meyer, 2015). Finally, attention to the unmentioned object follows a similar trajectory to the other-mentioned object for each participant. For the information-seeker, the unmentioned object is fixated most often in a similar time window as the answer object. For the information-giver, the unmentioned object is fixated most often in a similar time window as the question object. This suggests that the unmentioned object is considered to be a competitor for the other-mentioned object by both participants.

Combined, the first fixation onset time, gaze duration, and fixation pattern measures suggest that despite the relatively long total gaze durations in the experiment, the eye-movement data changes appropriately by condition and follows expected and systematic patterns across time. This provides support for the premise that gaze duration indexes the visual attention required for speaking and listening and suggests that it can serve as a suitable moderator variable for the relationship between memory and focus.

The relationship between visual attention and memory was quantitatively explored in two further models that included the total gaze durations spent on objects through the whole trial as a covariate. A visualization of these data appears in Figure 5. One of these models included gaze duration as a main effect only, and the other allowed it to interact with all other predictors. In both models, gaze duration was centered and scaled and the rest of the predictors were contrast-coded as described previously. The best-performing model was the one in which gaze duration was allowed to interact with the other predictors. In this model, gaze duration had a main effect on memory, such that more visual attention to objects overall was associated with better memory performance ($\beta = 0.58$, $SE = 0.07$, $z = 8.21$, 95% CIs [0.44, 0.72], $p < .001$). Importantly, gaze duration interacted reliably with the speaker benefit effect ($\beta = 0.28$, $SE = 0.10$, $z = 2.92$, 95% CIs [0.09, 0.47], $p < .05$), such that other-produced items (for the information-seeker = the answer; for the information-giver = the question) with short gaze durations were remembered less often than self-produced items with short gaze durations (for gaze durations under 500 ms, other-produced = 35% and self-produced = 63%). This meant that the speaker benefit effect was largest for objects with small gaze durations. In addition to these two significant effects, the effect of speaker role also became reliable in this model ($\beta = 0.20$, $SE = 0.07$, $z = 2.90$, 95% CIs [0.06,

Figure 4

Proportion of Fixations in Experiment 3 to Each Object Interest Area Out of Fixations to All Three Objects Across Time in Each Item Condition Split by Speaker Role



Note. Points are calculated as average across successive 100 ms time intervals with 95% confidence interval for each point. The top row is time-locked so that zero reflects the question noun onset, while the bottom row is time-locked so that zero reflects the answer noun onset. See the online article for the color version of this figure.

0.33], $p < .05$). This means that at the average object gaze duration (1,577 ms), information-seekers tended to remember the three objects on their screen better than information-givers did. Finally, the intercept in this new model was no longer significant, indicating that at the average object gaze duration, objects received even odds of "Yes" and "No" responses ($\beta = -0.08$, $SE = 0.09$, $z = -0.85$, 95% CIs [−0.25, 0.10], $p = .39$).

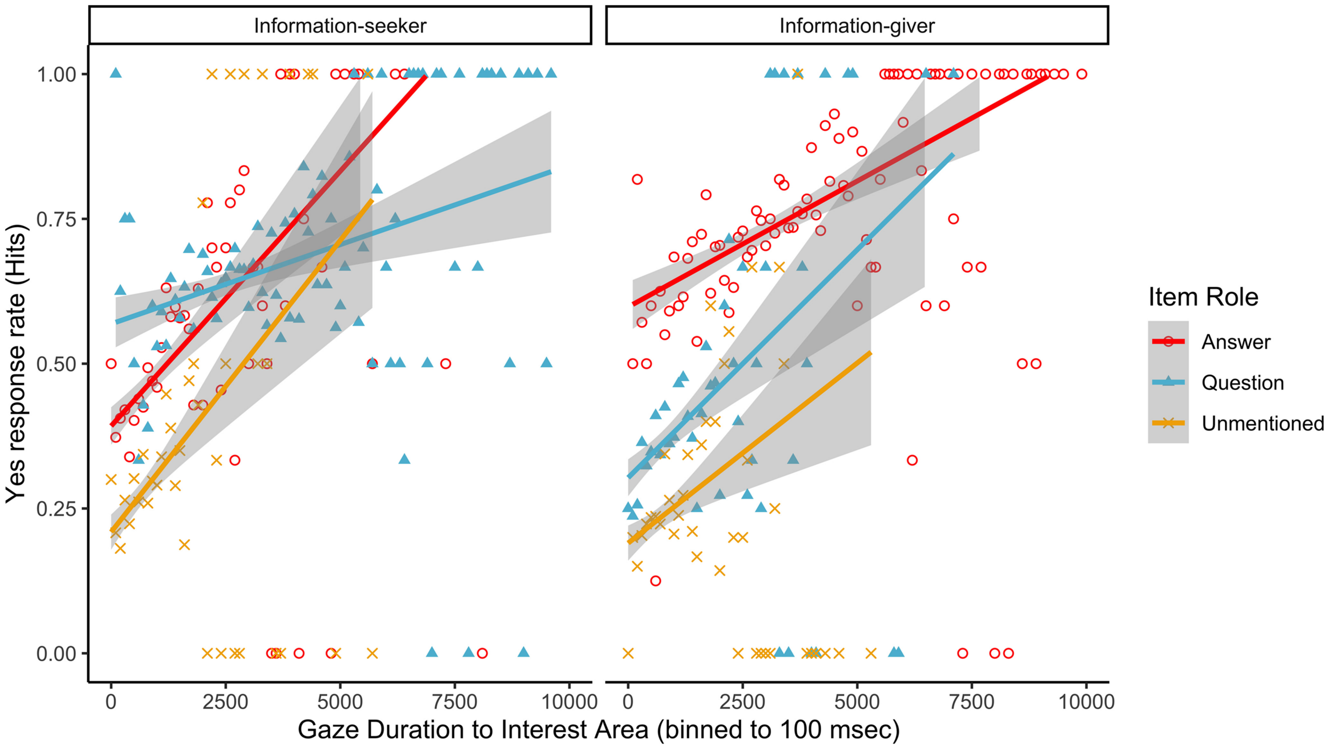
No other effects differed between this model and the model presented in Table 6. Most importantly, the main effect of production was still reliable in this model, as was the interaction between speaker role and production. This shows that the focus effects observed in this experiment are partly supported by visual attention, but that they remain present even when visual attention is accounted for.

Discussion

In this experiment, we investigated how speaker roles interacted with the speaker benefit effect in order to shed light on how focus affects memory in naturalistic conversations. We had participants

Figure 5

Hit Rates in the Experiment 3 Memory Task to Each Item Condition Split by Speaker Role and Total Gaze Duration to the Object During the Study Phase of the Experiment



Note. The information-seeker (in the left panel) asks questions and the information-giver (in the right panel) provides answers. Each colored point represents the average hit rate for all trials with the same gaze duration, binned to the nearest 100 ms. The colored lines reflect the best-fitting regression line for each item role, within each speaker role condition. For low values of gaze duration (left side of each panel), note that the other-produced item (information-seeker = answers, information-giver = questions) has a lower hit rate than the self-produced item, while for higher values of gaze duration, the hit rates become more equivalent. See the online article for the color version of this figure.

take turns producing utterances about objects that were similar to those used in Experiment 1, and then as in Experiments 1 and 2, we tested their memory for the names of those objects. The memory benefit for answers compared to questions was reliably modulated by the speaker benefit effect: individuals remembered the self-produced item better than the other-produced item, but information-seekers remembered the other-produced answers better than information-givers remembered the other-produced questions.

The overall benefit for self-produced speech over other-produced speech is consistent with earlier work on the generation effect and the production effect in one-participant production studies (Bertsch et al., 2007; Dew & Mulligan, 2008; MacLeod et al., 2010; Ozubko et al., 2014; Slamecka & Graf, 1978), and in recent studies of dialogue (e.g., Fischer et al., 2015; McKinley et al., 2017; Yoon et al., 2016). Overall, we showed that the speaker benefit effect remained remain large even in a conversational context, though the benefit for self-produced over other-produced speech was reliably reduced for items in focus. This suggests that it is important to consider interactions between predictors when generalizing to more naturalistic contexts.

Following the premise that questions put answers in focus, we explored the possibility that alternative answers (unmentioned items) would be remembered better by information-seekers than by information-givers. This was not the case. One possible explanation is that information-seekers did not deeply consider the unmentioned items as alternatives: they may have waited until they heard the referent, rather than trying to predict upcoming information. Inspection of the eye-tracking data suggests against this possibility though because for information-seekers, equal amounts of visual attention were directed to the answer and to the unmentioned item in the early portion of the trial (see Figure 4). It is therefore more likely that both speaker roles deemphasize the unmentioned item in memory.

We also explored the notion that visual attention directly supports both the speaker benefit effect and the focus effect. Answer objects were given more attention on average than question objects, and visual attention, indexed by gaze duration, was a reliable predictor of recognition memory and was particularly important in predicting memory for other-produced items. This shows how increased visual attention can overcome some of the memory issues associated with listening, compared to speaking, but that increased attention does not improve memory when the speaker benefit effect is already in play. The implication is that speaking and visual attention both improve memory for overlapping but nonidentical reasons. Visual attention also moderates the overall speaker role effect, so that when controlling for gaze duration, there was a memory benefit for information-seekers compared to information-givers. This suggests that visual attention is allocated by individuals based upon their speaker role, but that linguistic focus is distinct from visual attention. Combined, these findings show a rich and complex link between intention, attention, and the focus structure of a discourse.

General Discussion

Linguistic focus serves to draw a listener's attention to important elements of dialogues. In three experiments, we examined whether linguistic focus, therefore, created a memory benefit for nouns that served as answers compared to nouns in the questions that elicited them. Supporting our hypotheses, Experiment 1 showed a memory benefit for words appearing in short overheard dialogues that were used as answers compared to those used as questions. Experiment 2, where the same nouns were presented in two declarative sentences rather than in question–answer pairs, showed a significantly reduced memory benefit for the second-mentioned word when it appeared in a declarative sentence compared to in an answer, confirming the importance of linguistic focus in the results of Experiment 1. Experiment 3 examined the memory outcomes for words appearing in question–answer pairs produced in short conversations between two participants. Here, the goal was to show the interplay between focus and the general memory benefit for speaking over listening. Words used as answers again tended to be remembered better than those used as questions; this was evidenced by a larger difference between self-produced and other-produced items for people giving information than those seeking it. Combined, these findings show an important role for linguistic focus in individuals' resulting memory for language.

The clear and consistent advantage for answers compared to questions (Experiments 1 and 3) and compared to second items that were not in focus (Experiment 2) is important for the study of the memory representations that are developed during communication. This finding shows the consequences of linguistic focus: focus draws an interlocutor's attention to the items considered important, which then leads to memory benefits later on. These results are consistent with earlier work showing a memory benefit for focused compared to nonfocused items (e.g., Benatar & Clifton, 2014; Birch & Garnsey, 1995; Cutler & Fodor, 1979; Fraundorf et al., 2010; Johns et al., 2014; Sauermann et al., 2013; Sturt et al., 2004; Wang et al., 2011; Ward & Sturt, 2007; Yang et al., 2017). The current findings extend this earlier work to a within-item design: answers also receive a memory benefit compared to questions within the same dialogues. This confirms that focus effects on memory are isolable to the specific item in focus.

Experiment 3 also shows a clear replication of the speaker benefit effect shown in earlier work: a memory advantage for self-produced speech over other-produced speech even in a conversational context. This finding is attributable to a combination of two established effects in the memory for language literature. Generating labels for items, rather than reading or hearing them, creates a memory benefit known as the generation effect (Bertsch et al., 2007; Dew & Mulligan, 2008; Slamecka & Graf, 1978; Zormpa, Brehm, et al., 2019; Zormpa, Meyer, et al., 2019), while producing words aloud, rather than saying them silently, creates a memory benefit known as the production effect (MacLeod et al., 2010; Ozubko et al., 2014; Zormpa, Brehm, et al., 2019). In the elicited conversations in Experiment 3, individuals had to generate as well as produce their responses, enhancing memory due to both effects. This meant that following earlier work (e.g., Fischer et al., 2015; McKinley et al., 2017; Yoon et al., 2016; Zormpa, Brehm, et al., 2019), there was a general speaker benefit such that individuals tended to remember the item that they spoke about better than the other items. Importantly, our findings clearly demonstrate that both effects obtain in relatively simple conversations. Previous work has attributed the generation and production effects to increased item-specific processing (Hunt & McDaniel, 1993) or the increased distinctiveness of the resulting memory trace (Gardiner & Hampton, 1988; Ozubko et al., 2014). In other words, the speaker benefit effect likely arises because generation and production both require in-depth processing of the material to be produced, causing items to be encoded more deeply or with more detail. This deeper encoding occurs similarly in monologue and in at least some dialogue contexts.

In Experiment 3, we also found that focus moderated the speaker benefit effect. Other-produced words were remembered better by information-seekers than information-givers, meaning that when an item was in linguistic focus in the discourse, the speaker benefit effect was weaker. Importantly, the critical interaction between focus and the speaker benefit effect remained present even when visual attention (indexed by gaze duration) was accounted for. This suggests that part of what linguistic focus does is enhance the central processing of an in-focus item. We hypothesize that focus impacts memory via a similar mechanism to the speaker benefit effect: focus plausibly causes focused materials to be encoded more deeply and with more detail, leading to improved memory. Our hypothesis then is that both focus and speaking increase item-specific processing and/or distinctiveness, but that they are not identical phenomena. Future work examining the relationship between speaking and focus should be done in order to tease out the mechanisms and representations at play in each phenomenon.

Experiment 3 also tested the connection between gaze duration at encoding and later memory. Gaze duration, a measure of visual attention to the various objects in each discourse, was increased for focused items, and gaze duration was in turn reliably associated with memory performance. This replicates patterns shown in earlier work (Benatar & Clifton, 2014; Birch & Rayner, 1997; Lowder & Gordon, 2015). The speaker benefit effect was largest for items receiving relatively little visual attention. However, while gaze duration supported memory performance, it did not fully account for the speaker benefit and focus effects. Self-produced items with short gaze durations were still remembered quite well, and even when controlling for gaze duration, focus still affected the speaker benefit effect. Differences between patterns of visual attention and linguistic focus are important to consider in light of the difference between peripheral and central attention, which are often conflated in psycholinguistic research. While visual attention, which is a form of peripheral attention, is strongly associated with central attention (see Peterson et al., 2004), peripheral and central attention are disassociable under the right conditions (see, e.g., Johnston et al., 1995; Posner, 1980). Dissociations between visual attention and focus, like those shown in the current study, can therefore inform the field about how peripheral and central aspects of attention are used in language, how language requires integration of multiple sources of attention, and how higher-level cognition informs psycholinguistic processes.

Methodologically speaking, the results of these three experiments are important in highlighting similarities and differences between single-person and dyadic studies. The finding that speaker roles and focus influence memory for conversations highlights the importance of studying true dyads in psycholinguistics. Unlike natural conversations which generally involve two speakers, psycholinguistic studies of dialogue often involve one speaker who is responsible for "achieving" the goal of a trial, and a passive listener. This work demonstrates the need for studies where both (or all) speakers in a conversation have an active role to play: not only does this better approximate natural conversation in the lab, but the findings from Experiment 3 show that speaker roles have implications for what is remembered and produced during the experiment. It is an open question whether the results would generalize to a more complex experimental task (e.g., one in which objects actually need to be moved, sorted, or otherwise acted upon) or a more complex conversation (e.g., one that uses more complex utterances or a more complex discourse structure). We leave these as questions for future research.

The fact that qualitatively similar results obtain in Experiment 1 (passive listening) and Experiment 3 (active conversational participation) shows that linguistic focus in communication remains important regardless of whether there is an active conversational partner in the lab. However, quantitative differences in the size of the focus effect do emerge when comparing the two experiments directly. To test this question, we included experiment as a predictor in versions of the two analyses described in Experiment 1 that were run on the foils, unrelated items, and other-mentioned items in both experiments; experiment was contrast-coded as (E1: −0.5, E3: 0.5) and the analyses included random intercepts for item and participant, and random slopes by item and participant of the target–foil or question–answer contrast. The first model investigated whether the rate of "Yes" memory responses to targets and foils differed by experiment. In this model, there was a reliable main effect of the target–foil contrast and a reliable interaction between the target–foil contrast and experiment. The second model investigated whether there were differences in memory accuracy for questions versus answers across experiments. In this model, there was again a reliable main effect of the question–answer contrast and a reliable interaction between the question–answer contrast and experiment. Combined, these analyses show that foils (E1: 19%, E3: 19%) and questions (E1: 34%, E3: 34%) received identical rates of "Yes" responses across experiments, but that Experiment 3 showed a reduced rate of "Yes" responses for unmentioned items (E1: 30%, E3: 24%) and an increased rate of "Yes" responses for answers (E1: 40%, E3: 44%). The interactive context in Experiment 3, therefore, served to emphasize information that was most in focus (the answers) and deemphasize the information that was least in focus (the unmentioned items). Importantly, this penalty for unmentioned items holds in Experiment 3 despite the fact that the speaker benefit effect did not significantly modulate memory of the unmentioned items. Therefore, while linguistic focus matters even for overheard conversation, active participation as a listener or a speaker strengthens its effects. Dyadic experiments might be particularly well-suited then to examine smaller or more subtle effects of linguistic focus in the future.

Finally, the eye-tracking data in Experiment 3 show a potentially interesting point of comparison between dyadic and single-participant studies. Eye-tracking is frequently used in both language comprehension (Altmann & Kamide, 1999, 2007; Cooper, 1974) and production research (Ganushchak et al., 2014; Gleitman et al., 2007; Griffin & Bock, 2000; Konopka & Meyer, 2014; Van de Velde et al., 2014). The main purpose of collecting eye-tracking data in this study was to investigate the link between visual attention and linguistic focus, as discussed above. However, we also note that the temporal dynamics of visual attention, speaking, and listening (as highlighted in Figure 4) show compellingly similar patterns to those shown in single-participant studies: speakers attend to objects before producing their names, and listeners attend to objects after hearing their names. This suggests that eye-tracking two participants simultaneously in a conversation is a plausible method for future psycholinguistic research. Using this dyadic eye-tracking technique in varied experimental contexts could experimentally investigate the cognitive alignment between speakers in dialogues, testing some of the key premises of Pickering and Garrod (2004) or Dell and Chang (2014). This methodology might also prove useful to directly show whether the factors that enhance advance planning for speakers (for instance, Ganushchak et al., 2014; Gleitman et al., 2007; Konopka & Meyer, 2014; Van de Velde et al., 2014) also enhance prediction for listeners (for instance, Altmann & Kamide, 1999; Heyselaar et al., 2021; Hintz et al., 2017; Spivey et al., 2002).

We end with some comments about the utterances elicited in this study and a suggestion for further work. As we noted in the Introductary Part, we decided to use question–answer pairs to elicit focus because questions are far more common in spontaneous conversation than clefts, which have been often studied in the linguistic literature on focus. We suggest that future experimental work on memory for conversation should also be guided by consideration of the words and structures that actually appear with some frequency in spoken language. This means that the experimental work should go hand-in-hand with corpus analyses. Future work in this vein—that is, work studying experimentally the kind of language speakers use in everyday conversation—would situate basic research on memory and language within real-world conversation. It may, therefore, have direct implications for understanding how the specific language used affects social interactions, judgment, and decision making (see, e.g., Brown-Schmidt & Benjamin, 2018 for a discussion of memory for conversations in legal contexts.)

Conclusion

This work makes important advances to the understanding of what people remember from their conversations. Answers, which are in focus, were remembered better than questions that elicited them whether presented in an overheard conversation or in an active conversation. Speakers also remembered what they produced better than words produced by another speaker, and the difference between self-produced and other-produced information was greater when the participant was providing the answer rather than asking the question. Both patterns were supported by increased visual attention to objects that were used as answers and were self-produced, though visual attention did not account for the full memory benefit from focus or from speaking. Focus and the act of speaking, therefore, change what we attend to, what we encode, and what we later remember from conversations.

References

Almor, A., & Eimas, P. D. (2008). Focus and noun phrase anaphors in spoken language comprehension. *Language and Cognitive Processes*, *23*(2), 201–225. https://doi.org/10.1080/01690960701330936

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using LME4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benatar, A., & Clifton, C. (2014). Newness, givenness and discourse updating: Evidence from eye movements. *Journal of Memory and Language*, *71*(1), 1–16. https://doi.org/10.1016/j.jml.2013.10.003

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68. https://doi.org/10.1037/0096-3445.127.1.55

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Birch, S. L., Albrecht, J. E., & Myers, J. L. (2000). Syntactic focusing structures influence discourse processing. *Discourse Processes*, *30*(3), 285–304. https://doi.org/10.1207/S15326950dp3003_4

Birch, S. L., & Garnsey, S. M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*(2), 232–267. https://doi.org/10.1006/jmla.1995.1011

Birch, S. L., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *25*(5), 653–660. https://doi.org/10.3758/BF03211306

Birch, S. L., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition*, *38*(6), 740–752. https://doi.org/10.3758/MC.38.6.740

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, *142*(3), 791–808. https://doi.org/10.1037/a0029649

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, *5*(5), Article e10773. https://doi.org/10.1371/journal.pone.0010773

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. *PLoS ONE*, *9*(9), Article e106953. https://doi.org/10.1371/journal.pone.0106953

Brown, A. S., Jones, E. M., & Davis, T. L. (1995). Age differences in conversational source monitoring. *Psychology and Aging*, *10*(1), 111–122. https://doi.org/10.1037/0882-7974.10.1.111

Brown-Schmidt, S., & Benjamin, A. S. (2018). How we remember conversation: Implications in legal settings. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 187–194. https://doi.org/10.1177/2372732218786975

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In D. D. Steinberg & L. A. Jakobovits (Eds.), *An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 183–216). Cambridge University Press.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59. https://doi.org/10.1016/0010-0277(79)90010-6

Decuyper, C., Brysbaert, M., Brodeur, M., & Meyer, A. S. (2021). Bank of Standardized Stimuli (BOSS): Dutch names for 1400 photographs. *Journal of Cognition*, *4*(1), Article 30. https://doi.org/10.5334/joc.180

Dell, G. S., & Chang, F. (2014). The p-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), Article 20120394. https://doi.org/10.1098/rstb.2012.0394

Dew, I. T. Z., & Mulligan, N. W. (2008). The effects of generation on auditory implicit memory. *Memory & Cognition*, *36*(6), 1157–1167. https://doi.org/10.3758/MC.36.6.1157

Fischer, N. M., Schult, J. C., & Steffens, M. C. (2015). Source and destination memory in face-to-face interaction: A multinomial modeling approach. *Journal of Experimental Psychology: Applied*, *21*(2), 195–204. https://doi.org/10.1037/xap0000046

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357–383. https://doi.org/10.1016/j.jml.2006.07.004

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Ganushchak, L. Y., Konopka, A. E., & Chen, Y. (2014). What the eyes say about planning of focused referents during sentence formulation: A cross-linguistic investigation. *Frontiers in Psychology*, *5*, Article 1124. https://doi.org/10.3389/fpsyg.2014.01124

Gardiner, J. M., & Hampton, J. A. (1988). Item-specific processing and the generation effect: Support for a distinctiveness account. *The American Journal of Psychology*, *101*(4), 495–504. https://doi.org/10.2307/1423227

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Graesser, A. C., McMahen, C. L., & Johnson, B. K. (1994). Question asking and answering. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 517–538). Academic Press.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, *3*(2), 199–244. https://doi.org/10.1017/S0022226700016613

Heyselaar, E., Peeters, D., & Hagoort, P. (2021). Do we predict upcoming speech content in naturalistic environments? *Language, Cognition and Neuroscience*, *36*(4), 440–461. https://doi.org/10.1080/23273798.2020.1859568

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1352–1374.https://doi.org/10.1037/xlm0000388

Hjelmquist, E. (1984). Memory for conversations. *Discourse Processes*, *7*(3), 321–336. https://doi.org/10.1080/01638538409544595

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*(4), 421–445. https://doi.org/10.1006/jmla.1993.1023

Jarvella, R. J., & Collas, J. G. (1974). Memory for the intentions of sentences. *Memory & Cognition*, *2*(1), 185–188. https://doi.org/10.3758/BF03197513

Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: Tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience*, *29*(7), 877–892. https://doi.org/10.1080/01690965.2013.804941

Johns, C. L., Gordon, P. C., Long, D. L., & Swaab, T. Y. (2014). Memory availability and referential access. *Language, Cognition and Neuroscience*, *29*(1), 60–87. https://doi.org/10.1080/01690965.2012.733014

Johnston, J. C., McCann, R. S., & Remington, R. W. (1995). Chronometric evidence for two types of attention. *Psychological Science*, *6*(6), 365–369. https://doi.org/10.1111/j.1467-9280.1995.tb00527.x

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, *21*(6), 1590–1599. https://doi.org/10.3758/s13423-014-0620-7

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40. https://doi.org/10.1016/j.cogpsych.2014.04.001

Levelt, W. J. M. (Ed.) (1989). The generation of messages. In *Speaking: From intention to articulation* (pp. 107–160). MIT Press.

Lowder, M. W., & Gordon, P. C. (2015). Focus takes time: Structural effects on reading. *Psychonomic Bulletin & Review*, *22*(6), 1733–1738. https://doi.org/10.3758/s13423-015-0843-2

Machado, P., Romero, J., Nadal, M., Santos, A., Correia, J., & Carballal, A. (2015). Computerized measures of visual complexity. *Acta Psychologica*, *160*, 43–57. https://doi.org/10.1016/j.actpsy.2015.06.005

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition*, *45*(8), 1281–1294. https://doi.org/10.3758/s13421-017-0730-3

Miller, J. B. (1996). Memory for conversation. *Memory*, *4*(6), 615–632. https://doi.org/10.1080/741940999

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*(4), 465–501. https://doi.org/10.1016/0010-0285(78)90008-7

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE*, *7*(5), Article e37527. https://doi.org/10.1371/journal.pone.0037527

Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition*, *26*(6), 1313–1322. https://doi.org/10.3758/BF03201203

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition*, *21*(5), 689–698. https://doi.org/10.3758/BF03197199

Osaka, M., Nishizaki, Y., Komori, M., & Osaka, N. (2002). Effect of focus on verbal working memory: Critical role of the focus word in reading. *Memory & Cognition*, *30*(4), 562–571. https://doi.org/10.3758/BF03194957

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory*, *22*(5), 509–524. https://doi.org/10.1080/09658211.2013.800554

Peterson, M. S., Kramer, A. F., & Irwin, D. E. (2004). Covert shifts of attention precede involuntary eye movements. *Perception & Psychophysics*, *66*(3), 398–405. https://doi.org/10.3758/BF03194888

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, *17*(2), 132–138. https://doi.org/10.1080/17470216508416422

Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives* (Technical Report Cambridge NA Report NA2009/06). University of Cambridge.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

Sanford, A. J. S., Price, J., & Sanford, A. J. (2009). Enhancement and suppression effects resulting from information structuring in sentences. *Memory & Cognition*, *37*(6), 880–888. https://doi.org/10.3758/MC.37.6.880

Sauermann, A., Filik, R., & Paterson, K. B. (2013). Processing contextual and lexical cues to focus: Evidence from eye movements in reading. *Language and Cognitive Processes*, *28*(6), 875–903. https://doi.org/10.1080/01690965.2012.668197

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, *136*, 304–324. https://doi.org/10.1016/j.cognition.2014.10.008

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. https://doi.org/10.1016/S0010-0285(02)00503-0

Stafford, L., & Daly, J. A. (1984). Conversational memory. *Human Communication Research*, *10*(3), 379–402. https://doi.org/10.1111/hcre.1984.10.issue-3

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*(5), 882–888. https://doi.org/10.3758/BF03196716

van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, *39*(4), 973–978. https://doi.org/10.3758/BF03192992

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124–144. https://doi.org/10.1016/j.jml.2013.11.001

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813–820. https://doi.org/10.1016/j.neuropsychologia.2010.12.035

Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition*, *35*(1), 73–86. https://doi.org/10.3758/BF03195944

Yang, X., Zhang, X., Wang, C., Chang, R., & Li, W. (2017). The interplay between topic shift and focus in the dynamic construction of discourse representations. *Frontiers in Psychology*, *8*, Article 2184. https://doi.org/10.3389/fpsyg.2017.02184

Yoon, S. O., Benjamin, A. S., & Brown-Schmidt, S. (2016). The historical context in conversation: Lexical differentiation and memory for the discourse history. *Cognition*, *154*, 102–117. https://doi.org/10.1016/j.cognition.2016.05.011

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*(3), 340–352. https://doi.org/10.1080/09658211.2018.1510966

Zormpa, E., Meyer, A. S., & Brehm, L. E. (2019). Slow naming of pictures facilitates memory for their names. *Psychonomic Bulletin & Review*, *26*, 1675–1682. https://doi.org/10.3758/s13423-019-01620-x

Zormpa, E., Meyer, A. S., & Brehm, L. E. (2023a). *Answers are remembered better than the questions themselves* [Poster presented]. The Experimental Psychology Society (EPS) Meeting, Kent, Canterbury, United Kingdom. Retrieved from https://osf.io/x45ad

Zormpa, E., Meyer, A. S., & Brehm, L. E. (2023b). *Communicative intentions influence memory for conversations*. Retrieved from https://osf.io/y7seu

Received April 20, 2023
Revision received July 14, 2023
Accepted July 25, 2023 ■