Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–18

DOI: 10.1177/1747021819881265
qjep.sagepub.com
SAGE

Initiation of utterance planning in response to pre-recorded and "live" utterances

Matthias J Sjerps[1], Caitlin Decuyper[2] and Antje S Meyer[1,2]

Abstract

In everyday conversation, interlocutors often plan their utterances while listening to their conversational partners, thereby achieving short gaps between their turns. Important issues for current psycholinguistics are how interlocutors distribute their attention between listening and speech planning and how speech planning is timed relative to listening. Laboratory studies addressing these issues have used a variety of paradigms, some of which have involved using recorded speech to which participants responded, whereas others have involved interactions with confederates. This study investigated how this variation in the speech input affected the participants' timing of speech planning. In Experiment 1, participants responded to utterances produced by a confederate, who sat next to them and looked at the same screen. In Experiment 2, they responded to recorded utterances of the same confederate. Analyses of the participants' speech, their eye movements, and their performance in a concurrent tapping task showed that, compared with recorded speech, the presence of the confederate increased the processing load for the participants, but did not alter their global sentence planning strategy. These results have implications for the design of psycholinguistic experiments and theories of listening and speaking in dyadic settings.

Keywords

Conversation; turn-taking; utterance planning

Received: 9 November 2018; revised: 14 August 2019; accepted: 25 August 2019

Language is most commonly used in conversation (e.g., Levinson, 2016). Linguists and psycholinguists should therefore aim to understand the structure of conversations and the psychological processes occurring when people speak and listen to each other in everyday conversational contexts. Linguistic analyses of corpora of conversational speech have yielded important insights into the structural properties of conversation and led to hypotheses about the cognitive processes occurring in speakers and listeners engaged in conversation (Levinson & Torreira, 2015; Sacks, Schegloff, & Jefferson, 1974). However, testing these hypotheses through corpus analyses alone is challenging because researchers can only determine how speakers behaved under specific conditions, but cannot systematically vary these conditions. In our view, research into the processes underlying conversation therefore requires a two-pronged approach, involving the combination of corpus analyses and experimental research. This work contributes to the experimental research. Specifically, we asked how speech planning was affected by the presence or absence of an interlocutor. As we explain below, this is important for methodological reasons but also has implications for theories about speaking in dyadic contexts.

In conversation, people take turns, thereby switching roles as speakers and listeners (Levinson, 2016; Sacks et al., 1974; Stivers et al., 2009). Most of the time, only one person speaks, although there is substantial variation in utterance timing, periods of overlap (simultaneous talking) and gaps between turns tend to be short. As Levinson (2016) reports, periods of overlap have a modal duration of 100 ms, and the modal gap duration is around 200 ms (see also Heldner & Edlund, 2010; Stivers et al., 2009). This tight coupling between turns is remarkable because a 200-ms interval is too short for a speaker to plan even a

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Corresponding author:
Antje S Meyer, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands.
Email: antje.meyer@mpi.nl

short turn. Planning a content word, for instance, a picture name, takes at least 600 ms, and planning a short descriptive phrase can easily take more than a second (Indefrey & Levelt, 2004; Konopka, 2012). A 200-ms interval does, in fact, not even suffice to initiate a fully planned utterance (F. Ferreira, 1991; Piai, Roelofs, & Schriefers, 2011). Hence, to deliver their turn promptly after the end of the previous speaker's turn, upcoming speakers must already begin to plan their utterance during that turn. Moreover, they must predict when the current speaker's turn will end so that they can launch their utterance at the appropriate time (e.g., de Ruiter, Mitterer, & Enfield, 2006). This combination of processes—listening to the interlocutor, predicting their end of turn, and planning one's own utterance—renders turn-taking a complex cognitive task, in particular because both speech planning and listening require and compete for central processing resources (e.g., Cleland, Tamminen, Quinlan, & Gaskell, 2012; Kubose et al., 2006; Roelofs & Piai, 2011; Strayer & Johnston, 2001).

A working model of the way interlocutors may deal with this challenge has been proposed by Levinson and Torreira (2015). This model stipulates that listening and utterance planning occur largely in parallel. In everyday conversation, speakers can often identify the interlocutor's speech act (e.g., whether it is a request or comment) and the gist of the utterance early during the turn. It is proposed that based on this information, they immediately begin to plan their response to the level of the phonological form. This speech plan may be buffered until the interlocutor's turn is about to end. Then, the prepared utterance is launched. We will call the hypothesis that upcoming speakers plan their utterances as early as possible during the preceding turn the Early Planning Hypothesis.

One important research question arising from Levinson and Torreira's framework is how listeners identify imminent ends of turns (for further discussion and empirical work, see Bögels & Torreira, 2015; de Ruiter et al., 2006; Magyari, Bastiaansen, de Ruiter, & Levinson, 2014; Pickering & Garrod, 2004; Torreira, Bögels, & Levinson, 2015). A second important question, which is more directly relevant for this work, is when speakers initiate their speech planning, specifically whether they indeed begin to plan their utterances as early as possible. In the next section, we review the evidence relevant to this question.

Studies of utterance planning during listening

Several experimental studies have examined the Early Planning Hypothesis by asking participants to respond to spoken turns featuring response-relevant information either early or later during the turn and observing the effect of this manipulation on the response latencies and, in some cases, the participants' brain activity or their eye movements. Consistent with the Early Planning Hypothesis, all studies showed that the speakers began to prepare their utterances before the end of the preceding turn. Moreover, all studies, with the exception of our own earlier study (Sjerps & Meyer, 2015), showed that utterance planning began as early as possible during that turn. As a backdrop for this study, we briefly outline the method and relevant main results of each study.

The first study was carried out by Bögels, Magyari, and Levinson (2015). Participants worked together with a confederate who asked them quiz questions featuring early or late cues to the answer, as in "Which character, also called 007, appeared in the famous movies?" (Early Cue condition) or "Which character from the famous movies is also called 007?" (Late Cue condition). The average response latency was shorter by 310 ms in the Early than in the Late Cue condition, indicating that the participants began to plan their responses earlier when the cue to the answer ("007" in the example) appeared early than when it appeared late in the question. This conclusion was corroborated by analyses of the concurrently recorded electroencephalogram (EEG) signal, which suggested that the initiation of linguistic planning and a shift of attention from comprehension to production occurred approximately half a second after cue onset. These results were consistent with the Early Planning Hypothesis.

In a follow-up study aiming primarily to find further neurobiological evidence for early utterance planning, Bögels, Casillas, and Levinson (2018) used a different task. Here, the confederate asked the participants questions about objects they had just seen on their screen, for instance a banana and a pineapple. The questions were phrased such that the cue to the answer appeared early, as in "Which object is curved and is considered a type of fruit?" or late, as in "Which object is considered a type of fruit and is curved?". The results confirmed those of the earlier study: participants were substantially faster to respond in the Early Cue than in the Late Cue condition, and event-related potentials (ERPs) recorded time-locked to the onset of the cue ("curved") and corresponding non-informative word ("fruit") showed differences in positive effects, which the authors interpreted as evidence for response planning.

A similar paradigm was used in a behavioural study by Magyari, De Ruiter, and Levinson (2017). Here the questions were kept constant across conditions, but the displays varied such that response planning could begin early or late. For instance, when hearing the question "Which animal has a light-switch and a battery?", the participants could either see a display where each of two animals had some objects (Late Cue condition), or a display where one of the animals had no objects (Early Cue condition). Again, responses were given sooner after the end of the question in the Early than in the Late Cue condition.

Barthel and colleagues (Barthel, Meyer, & Levinson, 2017; Barthel, Sauppe, Levinson, & Meyer, 2016) used an

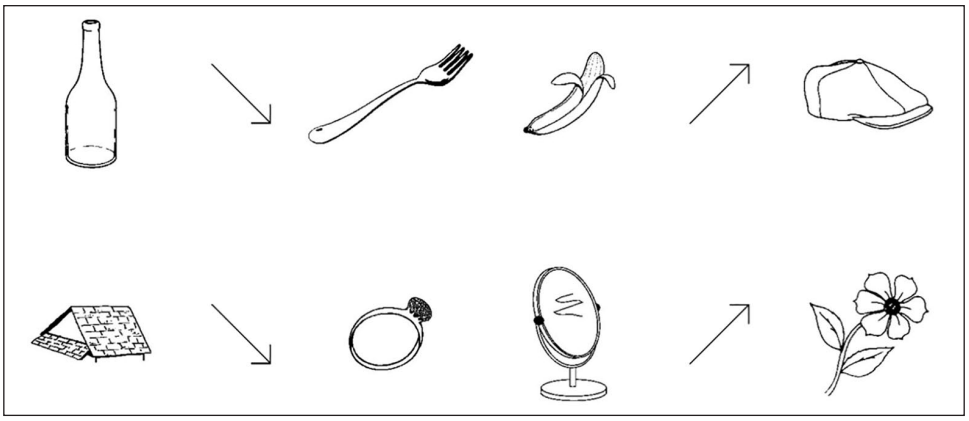

Figure 1. Example display used in both experiments.

interactive sentence-completion paradigm, where a confederate named objects she could see on her screen and the participant completed the description by describing any additional objects presented on their screen. The experiment was conducted in German, and the confederate's utterances ended either in a noun or a verb form, as in "Ich habe einen Schluessel und einen Stift (besorgt)" ("I have a key and pen (acquired)."). The participants produced utterances such as "Ich habe eine Melone und eine Zeitung" ("I have a melon and a newspaper."). In addition to the utterance latencies, measured from the end of the confederate's turn, the participants' eye movements were recorded. The speech onset latencies were shorter when the turn ended in a verb form than when it ended in a noun, indicating that utterance planning began when the final noun phrase, rather than the end of the turn, was perceived. The analyses of the participants' eye movements confirmed this conclusion: the frequency of the participants' looks to the first object they mentioned rose sharply after the onset of the last noun phrase in the confederate's description, regardless of whether or not a verb form followed.

In two further studies, participants were asked to answer polar (yes/no) questions. In two speech planning experiments reported by Corps, Crossley, Gambi, and Pickering (2018), participants heard recorded questions with highly predictable endings such as "Are dogs your favourite animals?" or with less predictable endings, such as "Have you ever injured your eye?". Participants were instructed to respond as quickly as possible. They were faster to do so in the predictable than in the unpredictable condition, indicating that they began to plan their utterance before the end of the question. In a study by Meyer, Alday, Decuyper, and Knudsen (2018), participants answered polar questions such "Do you have a green sweater?" about objects on their screen. The questions were kept constant across conditions, but the displays varied: in the Early Cue condition, the four objects on the screen had the same colour, and participants could therefore respond as soon as they had processed the colour adjective. In the Late Cue condition, the objects were shown in different colours, and the participants could therefore only respond after they had processed the noun. Consistent with the results found in the studies reported above, responses were given sooner after question offset in the Early than in the Late Cue condition.

The final relevant study was carried out by Sjerps and Meyer (2015). Since this research is closely related to the present study, its design and findings are described in detail. Dutch participants saw displays featuring two rows of objects (see Figure 1). There were three conditions. In the Speaking-Only condition, participants first heard a recorded utterance referring to the objects shown in one row, for instance, the Dutch translation equivalent of "Put the bottle below the fork and the banana above the hat." The arrows between the objects indicated which prepositions should be used. Immediately following the recording, the participants should provide a description of the other row using the same utterance format, saying, for instance, "Put the roof below the ring and the mirror above the flower." In the Tapping and Speaking condition, participants performed the same task, but in addition carried out a complex finger tapping task throughout the block of test trials. Finally, in the Listening-Only condition, the participants carried out the tapping task while listening to recorded descriptions of both rows of objects delivered by different speakers. In all three conditions, they had to monitor the second utterance of each pair for correctness. Depending on the condition, this was their own utterance or the pre-recorded utterance of another speaker. Tapping performance was recorded as an indicator of the mental load arising during the task (e.g., Boiteau, Malone, Peters, & Almor, 2014; Kemper, Herman, & Lian, 2003; Kemper, Herman, & Nartowicz, 2005; Somberg & Salthouse, 1982). It was expected to be better (i.e., show more correct taps per second; for details, see the section "Tapping Performance" below) in a baseline period between trials, when the tapping task was carried out by itself, than while participants were listening to the recorded utterances or speaking themselves. This is because both listening to

speech and speaking require processing capacity and should therefore interfere with the tapping performance (e.g., Cleland et al., 2012; Cook & Meyer, 2008; V. S. Ferreira & Pashler, 2002; Mattys, Brooks, & Cooke, 2009). A second prediction, also based on earlier dual-task studies, was that tapping performance would be worse during speaking than during listening. This is because speech planning and speaking have been shown to require more cognitive capacity than listening to speech (e.g., Almor, 2008; Kunar, Carter, Cohen, & Horowitz, 2008; Recarte & Nunes, 2003). Additional interference between tapping and speaking may arise because both tasks involve multiple response selection processes (e.g., Hegarty, Shah, & Miyake, 2000). The most important question was when the deterioration of tapping related to speech planning and speaking would occur. In a study by Boiteau and colleagues (2014), participants carried out a visuomotor task (tracking a moving target on their screen using the mouse) while engaged in free conversation. As expected, the authors found poorer tracking performance when participants were speaking than when they were listening to their partner. Importantly, the deterioration of tracking performance already began well before speech onset, towards the end of the partner's turn, suggesting that it was related to the onset of speech planning. Based on these results, Sjerps and Meyer predicted that in their study tapping performance should already deteriorate before the offset of the preceding utterance.

In one of the two experiments reported by Sjerps and Meyer, the participants' eye movements were recorded in addition to their speech and tapping performance. Earlier studies had shown that people listening to utterances referring to the visual environment tended to fixate the relevant locations (e.g., Cooper, 1974; Huettig, Rommers, & Meyer, 2011) and that, similarly, people naming or describing several objects tended to look at each of them around 500 ms to a second before mentioning them (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998). This holds even when the speakers are familiar with the objects and can identify them without fixating upon them (e.g., Meyer, Wheeldon, Van der Meulen, & Konopka, 2012). Moreover, the durations of the gazes to the objects to be named have been shown to depend on the time needed to recognise the objects as well as the time needed for the retrieval of their names (e.g., Griffin & Bock, 2000; Jongman, Roelofs, & Meyer, 2015; Meyer & Van der Meulen, 2000). A likely basis for the strong link between eye gaze and speech processing is that a person's point of gaze is tightly linked to the focus of their visual attention and that directing one's attention to an object greatly facilitates processing it (e.g., Irwin & Gordon, 1998; Rommers, Meyer, & Praamstra, 2017). Sjerps and Meyer therefore expected that the participants would first look at the objects mentioned in the recorded utterance and would turn to their own objects when they began to plan the utterances about them.

The most important question for the study was when this shift of gaze towards the participant's objects and the deterioration of tapping performance would occur. Whether the recorded description referred to the top or bottom row varied randomly from trial to trial, but participants knew which row was being described as soon as they had heard the first object noun and could then begin to plan their own utterance. If the Early Planning Hypothesis holds, evidence for utterance planning should be seen very early during the trial.

As expected, the participants' tapping performance changed across the course of the trials: compared with the baseline period before the onset of the recorded utterance, the tapping performance deteriorated slightly while the participants were listening to the description of the first row of objects, but was much poorer when they were speaking themselves. Importantly, the steep decline in tapping performance did not begin early during the recorded utterance, but only between 1 and 0.5 s before its end. The analysis of the participants' eye movements showed that they shifted the focus of their attention from the objects mentioned in the recording to the objects they were about to describe themselves around the same time, shortly before they took over the turn. Hence, both the tapping performance and the gaze behaviour suggested that speakers initiated their utterance planning before the end of the recorded description. However, they did not do so at earliest opportunity, as predicted by the Early Planning hypothesis, but only just before the anticipated end of the previous turn.

In sum, all studies of speech planning during listening clearly indicate that upcoming speakers do not await the end of the preceding turn, but already begin to plan their utterance during that turn. However, speech planning does not necessarily begin at the earliest possible moment. Speakers appear to have some flexibility in the way they combine listening and speech planning: Sometimes, they begin to plan a response to a preceding utterance as soon as the response-relevant information is available (e.g., Bögels et al., 2018; Bögels, Magyari, & Levinson, 2015), but on other occasions, they postpone planning until shortly before the end of the preceding turn (Sjerps & Meyer, 2015).

This conclusion leads to the question which variables affect how speakers combine speaking and listening. Many social, pragmatic, and linguistic factors that might bias speakers towards early or late speech planning come to mind. The two studies that, taken together, illustrate the speakers' flexibility in speech planning most clearly are those conducted by Bögels, Magyari, and Levinson (2015) and by Sjerps and Meyer (2015). They differ in many features. In the comparison of these studies, Bögels and colleagues (2018) highlight that their paradigm was "truly interactive" (p. 296), as the participants believed they interacted with a confederate and answered their

questions. By contrast, the paradigm used by Sjerps and Meyer,

> deviated quite far from typical conversation interaction in that participants knew that they were talking to a pre-recorded voice and that there was no contingency between the spoken content of their [own] and the computer's turns. In other words, participants were just alternating their speech with the computer's without the computer's speech having any bearing on their own speech plans and vice versa. (p. 296)

One should note that in Sjerps and Meyer's paradigm, the pre-recorded utterances and the participants' utterances were linked (contrary to the above claim) as the participants had to describe the set of objects not covered in the recording and that the participants in the study by Bögels and colleagues sat in an experimental booth and could only hear the confederate via headphones, which also deviates quite far from typical conversation. Nevertheless, the utterance pairs were probably linked in a more natural way in the study by Bögels and colleagues. Moreover, the set-up resembled a party game or pub quiz, and perhaps the similarity to such situations encouraged participants to respond fast and plan their utterances early. In addition to the communicative situation, the two studies differed in the linguistic structure of the spoken input. Bögels and colleagues used questions that varied in length and structure from trial to trial, whereas the recorded utterances in Sjerps and Meyer's study all had the same structure and length. This consistency of the input and the predictability of the end of turn may have encouraged the participants to align their speech planning with a particular "landmark," such as the onset of the last noun phrase, in the recorded description. Finally, the participants' responses in the two studies were also different; they were short phrases (e.g., "James Bond" or "Paris") in the study by Bögels and colleagues and multi-phrase sentences in the study by Sjerps and Meyer. As Sjerps and Meyer argued, preparing these phrases early would probably have led to substantial interference between the speech input and the speech plan. In addition, all utterances began with the same two words, "zet de . . ." ("put the . . ."), which provided the participants with planning time for the following nouns during the articulation of these two words. Thus, in this study, late planning was an effective strategy, allowing participants to respond quickly after the end of the interlocutors' turn while minimising their memory load.

The comparison between these studies illustrates the multitude of variables that may affect the timing of utterance planning. Further empirical work is needed to determine which variables actually have an impact on the timing of speech planning in dyadic settings and through which cognitive mechanisms they affect the speakers' behaviour.

This study

In this study, we returned to the paradigm used in the study by Sjerps and Meyer and compared the participants' speech planning when they responded to utterances produced "live" by a confederate, who set next to them and looked at the same screen, or to recorded utterances from the same speaker. Examining the effects of confederate presence was considered to be important for methodological reasons, specifically for the design of future studies into speech planning in dyadic situations. As noted above, a number of authors have suggested that the presence or absence of an interlocutor may strongly impact participants' speech planning. Investigating this variable is also important for theoretical reasons. If substantial effects of confederate presence are seen, further studies can be directed at understanding the bases of these effects and thereby help to link experimental psycholinguistics and social psychology and social neuroscience.

In the existing studies on the onset of speech planning, the effect of interlocutor presence or absence has not been assessed. In the studies by Bögels and colleagues (2018; Bögels, Magyari, & Levinson, 2015) and by Barthel and colleagues (2016; Barthel et al., 2017), participants were tested in sound-attenuated booths and heard input they believed to be produced live by a confederate seated outside of the booth. In the study by Magyari and colleagues (2014), Corps and colleagues (2018), and Sjerps and Meyer (2015), they were informed that they would hear and respond to recorded utterances. In the study by Meyer and colleagues (2018), participants answered polar (yes/no) questions, which were either pre-recorded or posed by another participant. Analyses of the speech onset latencies showed no marked differences between the experiments, but due to differences in their designs and procedures, no strong conclusions about the effects of the presence of an interlocutor could be drawn. In short, it is unknown whether speakers time their spoken utterances differently when they respond to pre-recorded utterances or utterances produced "live" by a confederate.

Studies using other paradigms have not provided directly relevant information either. In fact, surprisingly, little is known about the effects of the presence or absence of an interlocutor on speech planning and the timing of utterances. Earlier behavioural studies have compared the content of utterances produced in monologue and dialogue (e.g., Kantola & van Gompel, 2016; Murfitt & McAllister, 2001; Tolins, Zeamer, & Fox Tree, 2018), and recent neurobiological studies have begun to explore the brain circuits specifically activated in social contexts (e.g., Kuhlen, Bogler, Brennan, & Haynes, 2017; Sassa et al., 2007; Schilbach et al., 2013), but none of these studies specifically investigated the timing of speech planning. Swets, Jacovina, and Gerrig (2013) directly compared utterance planning in the presence versus absence of a listener. In this study, participants described complex tangram figures. The authors found that participants described the figures more efficiently when a listener was present, that is, provided more information in about the same amount of time,

than in the absence of a listener. However, this study did not analyse gaps between turns.

As noted, Bögels and colleagues (2018) and Barthel and colleagues (2016) have argued that the (assumed) presence of an interlocutor may be important for eliciting natural turn-taking behaviour, and specifically early utterance planning. One reason for this might be that in natural conversations, unexpectedly long pauses before a response may carry pragmatic meaning, for instance, signalling reluctance to respond (e.g., Bögels, Kendrick, & Levinson, 2015; Kendrick & Torreira, 2015). If this hypothesis is correct, utterance preparation should begin earlier in the presence than in the absence of an interlocutor. The paradigm used by Sjerps and Meyer was well-suited to test this hypothesis because in the original study, very late utterance planning was observed while much earlier planning would have been possible. Thus, there is ample opportunity to observe a shift from late to early planning, should it occur. If participants opt for earlier utterance planning in the presence than in the absence of a confederate, their speech onset latencies should be shorter than observed before, and one should see earlier fixations to the participants' own objects and an earlier decline in tapping performance relative to the baseline.

The experiments reported below were similar to the dual-task and eye tracking experiment (Experiment 2) by Sjerps and Meyer (2015). In Experiment 1, the same materials were used and participants again first heard the description of one row of objects and then provided a description of the other row. Participants either performed only this task (Speaking-Only condition) or also carried out the continuous tapping task (Tapping and Speaking condition). The most important difference to the earlier study was that the first description was now delivered by a confederate, who sat next to the participant and looked at the same screen.

As before, we determined the gap durations (i.e., the participants' speech onset latencies measured from the offset of the confederate's turn) and assessed the timing of the participants' speech planning through analyses of their eye movements and tapping performance. If participants employ the same planning strategy as before, they should fixate upon the objects named by the confederate until about a second before the end of the confederate's utterance. In the Tapping and Speaking condition, their tapping performance should deteriorate around the same time. By contrast, if the participants initiate their utterance planning earlier than before, one should see earlier fixations to the participant's own objects and an earlier decline in tapping performance. In an extreme case, utterance planning could begin as soon as the pictures appeared on the screen, that is, about 2 s earlier than observed before.

Although Experiment 1 of this study was similar to Experiment 2 reported by Sjerps and Meyer (2015), a direct comparison of the results was not appropriate because there were some differences in their designs. Specifically, in this study, we did not include the Listening-Only condition again, but only used the Speaking-Only and Tapping and Speaking conditions. Furthermore, we did not ask participants to evaluate the correctness of the utterances, which should make the task more similar to everyday communication tasks. Finally, rather than sometimes describing the top and sometimes the bottom row of objects, the participants now always described the bottom row. This change in the design was introduced to facilitate the interpretation of the eye movements. Whereas in Sjerps and Meyer's (2015) study looks to objects early in the trial could result from uncertainty about the row being described, now participants knew that the confederate would always describe the top row and that they would have to describe the bottom row. Consequently, it was now more likely that looks to the top row indicated attentive listening, whereas looks to the bottom row indicated preparation for speaking. Because of these differences between the present Experiment 1 and the experiments described by Sjerps and Meyer (2015), this study includes a second experiment involving pre-recorded speech. In that experiment, the participants heard recordings of utterances made by the confederate during Experiment 1. Thus, we could directly compare the participants' speech planning when they listened to utterances produced live by the confederate or to recorded utterances of the same speaker.

As Experiment 2 was conducted after Experiment 1 and participants were not randomly assigned to the two experiments, it is, strictly speaking, not appropriate to treat the two experiments as conditions of one study. However, as the methods of the two experiments, the characteristics of the samples, and the results were very similar, we describe the methods, analyses, and experimental results of the two experiments together.

Method

Participants

The experiments were carried out with 69 paid participants. Data of 13 participants of Experiment 2 had to be discarded. One of these participants did not complete the experiments; in one session, the button box was not correctly connected to the computer and no data were recorded, and in the remaining sessions, the eye movement data were not of sufficient quality (i.e., the calibration procedure failed) in one or more blocks. This left data of 27 participants from Experiment 1 (mean age = 21.21 years, $SD = 1.79$) and data of 29 participants from Experiment 2 (mean age = 21.95 years, $SD = 1.98$) for the analyses. All participants were female to match the gender of the confederate. They were university students, native speakers of Dutch, and reported normal hearing and normal or corrected to normal vision. Ethical approval for the study had

been granted by the Ethics Board of the Social Sciences Faculty of Radboud University, Nijmegen.

Apparatus

A remote Eyelink 1000 eye-tracker (SR Research) and the software packages Experiment Builder and Data Viewer (SR Research) were used to control the experiment and record and convert the participants' eye movement data. The movements of the left eye were recorded with a sampling frequency of 500 Hz. Finger tapping was recorded by means of a purpose-built four-button box attached to the computer. The buttons were microphones that required very little pressure to record a response and only produced a barely audible sound upon tapping. The participants' speech was recorded using a Sennheiser ME 64 microphone. Praat software (Boersma & Weenink, 2018) was used to measure the onsets and durations of the confederate's and participants' utterances. In Experiment 2, auditory stimuli were presented using Sennheiser HD 201 headphones.

Materials

Visual stimuli. The same visual stimuli were used as in our earlier study (Sjerps & Meyer, 2015). Forty black-on-white line drawings had been selected from the picture database generated by Severens, Van Lommel, Ratinckx, and Hartsuiker (2005). The names of the selected pictures were monomorphemic and monosyllabic or disyllabic. The log frequency and the log naming latency of all picture names were within 1.5 standard deviations of the corresponding averages for all pictures in the database.

Using these pictures, five practice displays and 90 experimental displays were created. Each display featured eight objects, randomly arranged in four pairs on two rows as shown in Figure 1. There were no strong semantic or phonological relationships between the items in a display. Each object appeared 18 times on experimental trials (nine times per condition, see below) and twice on practice trials (once per condition).

In each row, arrows pointing up or down appeared between the first and second object and between the third and fourth object. Each display featured two arrows of each type, assigned randomly to the four positions. The line drawings were sized to have a maximal length and width of 7 cm, corresponding to a visual angle of approximately 6.5° for the participant. The length of the arrow was 4.5 cm, corresponding to approximately 4°. The distance between the top and bottom row of objects (object midpoint-to-midpoint) was 10.5 cm, corresponding to approximately 9.8°. Thus, there was a blank space of approximately 3.5 cm between the top and bottom row of objects. With effort, participants could probably garner some information about objects in the bottom row while fixating objects in the top row (please see below for further discussion).

Auditory stimuli. In Experiment 1, the auditory stimuli were produced by the confederate (female, 24 years). Upon presentation of the display, she described the top row of pictures, saying, for instance, "Zet de fles onder de vork en zet de banaan boven de pet" ("Put the bottle below the fork and put the banana above the hat"). The participant then described the bottom row using the same utterance format. The confederate was trained to use a consistent moderate speech rate, but naturally there was some variation in her speech rate, and she occasionally committed speech errors. Offline analyses of her speech showed that the average onset time of her utterances was 594 ms ($SD = 164$), and the average duration of her utterances was 2.91 s ($SD = 0.28$). Naming errors occurred on 0.8% and 0.6% of the trials in the Tapping and Speaking and the Speaking-Only condition, respectively, and noticeable hesitations occurred on 16.2% and 15.7% in the Tapping and Speaking and Speaking-Only condition, respectively. Given that speakers in everyday conversations are also sometimes disfluent and repair themselves, these trials were not excluded from the analyses. 0.3% of the trials were excluded due to failure of the equipment.

To create the input for the participants in Experiment 2, we selected the confederate's recordings from one session of Experiment 1. In this session, the average turn onset latency (631 ms, $SD = 152$) and duration (2.81 s, $SD = 0.15$) were very similar to the average across all sessions, and there were 10 hesitations and no errors. Since participants may expect recordings to be flawless, we replaced the 10 trials with hesitations by recordings from another session.

Design

Within each experiment, there were two within-participants experimental conditions: Tapping and Speaking and Speaking Only. Each condition was tested in a separate block. The order of the blocks was counterbalanced across participants. The 90 experimental displays were randomly assigned to two sets of 45 displays each, and the assignment of sets to conditions was counterbalanced across participants. The experimental displays within blocks were presented in a random order. Each block began with five practice displays. Four experimental lists were created (two presentation orders of conditions crossed with two assignments of sets of pictures to conditions).

Procedure

In Experiment 1, the experimenter introduced the confederate to the participant and explained that the confederate would "join in with the task." Thus, participants were not deceived into thinking that the confederate was another

naive participant. The participants were asked to study a booklet showing the objects used in the experiment along with the object names. They were asked to use these names in the description task. Then, they performed a tap training to familiarise themselves with the button box and the tapping sequence. Participants tapped a complex pattern of 1 (index finger), 3 (ring finger), 2 (middle finger), and 4 (little finger) with their right hand. The experimenter monitored the tapping rate indicated on his monitor and encouraged the participants to tap faster when the rate fell below three taps per second. The training was terminated as soon as the participant had performed 50 consecutive correct taps. Training times varied across participants between 2 and 10 min.

Next, the participants received another booklet showing the 40 pictures, now without their names, and were asked to name them. Naming errors were immediately corrected by the experimenter. Then, the participants were instructed for the first block of the main experiment. They were told that they would carry out a description task together with the confederate. The experimenter showed them a screenshot of a visual display and explained that on every trial, the confederate would first describe the objects in the top row and the participant should then describe the objects in the bottom row using the same utterance format. They should try to complete the utterance before the end of the trial, which was signalled by the disappearance of the visual display. In the Speaking-Only condition, no further instructions were given. In the Tapping and Speaking condition, participants were additionally asked to tap the trained pattern as quickly as possible throughout the test blocks. They did not tap in the breaks between blocks.

Following the instructions, the participants were seated in front of the eye-tracker, approximately 1 m away from the screen. Head movements were unrestricted but participants were asked to keep their eyes on the screen throughout the trial. The eye-tracker was calibrated before each test block. Intermediate validations were performed after every 15 trials. The confederate sat next to the participant and viewed the same screen.

Each trial began with a blank interval of 4 s, followed by the presentation of the display. The display remained on screen for 8 s, followed by a 1-s blank screen. Then the next trial began. In Experiment 1, the confederate described the top row of objects and the participant the bottom row. In Experiment 2, the participants heard a recording of the confederate. In both experiments, participants were told that they would hear the description of the first row of objects and should provide the description of the second row using the same utterance format. Each of the two test blocks included 45 experimental trials and took approximately 15 min to complete.

Analyses and results

The participants' utterances were transcribed, with errors and hesitations being noted, by a trained native speaker. Speech onsets and offsets were measured manually by trained native speakers of Dutch using the software package Praat (Boersma & Weenink, 2018). Statistical analyses were carried out by fitting models with the lmer() function in the lme4 package (Bates, Maechler, & Bolker, 2013) in *R* (R Core Team, 2013). The optimal model among them was chosen based on model comparisons with the anova() function, which performs a likelihood ratio test for quality of fit. Unless specified otherwise, categorical variables were sum-to-zero contrast-coded (for Experiment: Exp1 = 1, Exp2 = −1; for Task: Speaking Only = 1, Tapping and Speaking = of −1; for Task Order: Speaking Only First = 1, Tapping and Speaking First = −1).

Speech errors and hesitations

As expected, the participants' speech was different in the Tapping and Speaking condition compared with the Speaking-Only condition. The differences mostly consisted of noticeable hesitations. Hesitations occurred on 24.7% of the Tapping and Speaking trials and 10.4% of the Speaking-Only trials of Experiment 1, and on 28.0% of the Tapping and Speaking and 13.0% of the Speaking-Only trials of Experiment 2. Occasionally, participants used incorrect object names. This happened on 2.0% of the Tapping and Speaking trials and on 1.2% of the Speaking-Only trials in Experiment 1, and on 2.9% of Tapping and Speaking and 2.2% of the Speaking-Only trials in Experiment 2. In addition, participants sometimes failed to complete their description before the end of the trial (i.e., within 9 s after picture onset). Such timeouts were recorded on 3.4% of the Tapping and Speaking trials and on 0.8% of the Speaking-Only trials in Experiment 1, and on 0.9% of Tapping and Speaking trials and on 0.5% of Speaking-Only trials in Experiment 2. Typically, timeouts arose because the participant was unable to retrieve the name of a picture (e.g., "put the euuh . . . p . . . eeuh . . . [end of trial]"), not because the confederate completed the utterance exceptionally late. Note that several error types could occur on the same trial. Subjectively fluent (to the annotator) and correct utterances occurred on 72.4% of the Tapping and Speaking trials and 87.8% of the Speaking-Only trials of Experiment 1, and on 70.4% of the Tapping and Speaking and 85.7% of the Speaking-Only trials of Experiment 2.

A binomial model with the dependent variable Fluency (0: *non-fluent* vs 1: *fluent*) was used to statistically assess differences in fluency across conditions. The optimal model included the fixed effects Experiment, Task, and Task Order. The random effect structure of this model consisted of intercepts and slopes for the factor Task for Participants and intercepts for Displays. The model ($\beta_{Intercept} = 1.55$, $z = 16.32$, $p < .001$) revealed only a significant main effect of Task ($\beta = 0.49$, $z = 10.87$, $p < .001$). The positive β estimate for the effect of Task indicates that more fluent

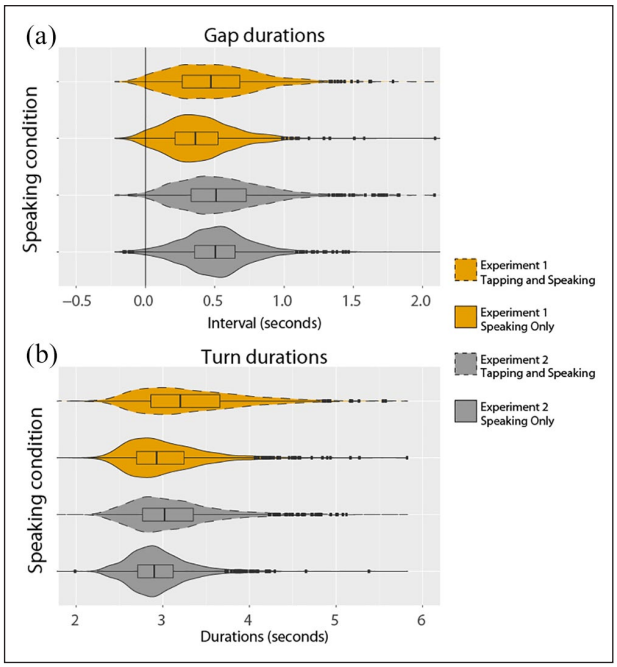

Figure 2. (a) Distributions of gap durations and (b) participants' turn durations in the Tapping and Speaking and in the Speaking-Only conditions of Experiment 1 (live confederate) and Experiment 2 (pre-recorded confederate). Boxplots indicate median and 25% and 75% quantiles per condition.

descriptions were given in the Speaking-Only task than in the Tapping and Speaking task. The model also showed a main effect of Task Order ($\beta = 0.24$, $z = 2.75$, $p < .01$), suggesting that participants who started with the Speaking-Only task were more fluent than participants who started with the Tapping and Speaking task. There was no main effect of Experiment. Thus, the presence or absence of an interlocutor did not influence participants' fluency.

Gap and turn durations

For the analysis of the gap durations (i.e., the participants' speech onset latencies, measured from the offset of the confederate's turns), trials that featured naming errors or timeouts were excluded (3.7% and 3.1% of the trials of Experiment 1 and 2, respectively). Figure 2a shows the distribution of gap durations per condition in the two experiments as violin plots. Time point zero is the offset of the confederate's turn (the end of the final phoneme). Negative gap durations occurred when participants began to speak before the confederate had completed her utterance. For the analyses, values deviating from the grand mean per experiment by more than 2.5 *SD* were considered outliers and excluded (1.5% of the trials in Experiment 1 and 1.9% of the trials in Experiment 2).

To analyse the gap durations, a statistical model was used that included the fixed effects Experiment and Task and the interaction between these effects. Task Order was added as a covariate. The random effect structure consisted of intercepts and slopes for the effects of Task on Participants and intercepts for Displays. The optimal model ($\beta_{Intercept} = 0.49$, $t = 19.86$, $p < .001$) included a non-significant main effect for Experiment ($\beta = –0.04$, $t = −1.72$, $p = .086$), a main effect of Task ($\beta = –0.04$, $t = −3.91$, $p < .001$), a non-significant main effect for Task Order ($\beta = 0.01$, $t = 0.55$, $p = .58$), and a non-significant interaction between Task and Experiment ($\beta = –0.02$, $t = −1.72$, $p = .086$). The negative β estimate for Task shows that participants were overall faster to take over the turn when only speaking than when tapping and speaking. The close to significant effect of Experiment and the interaction between Task and Experiment reflect a tendency for responses to be slower in general and the difference between the tasks to be smaller in Experiment 2 (using recorded utterances) than in Experiment 1 (using live utterances).

The average durations of the participants' turns are displayed in Figure 2b, measured from the onset of the verb "Zet" to the end of the last noun. The same statistical analysis was carried out as for the gap durations. 2.3% of the trials of Experiment 1 and 2.6% of the trials of Experiment 2 were excluded as outliers (durations beyond 2.5 *SD* away from the mean). The optimal model included the dependent variable Duration, the fixed effects Experiment and Task and their interaction, and a random effect structure with intercepts and slopes for Task for Participants and intercepts for Display. Adding Task Order as a covariate did not improve the model fit. The optimal model ($\beta_{Intercept} = 3.10$, $t = 88.78$, $p < .001$) revealed a main effect of Task ($\beta = –0.12$, $t = –8.24$, $p < .001$), reflecting the longer duration of participants' turns when they were engaged in the tapping task. In addition, a significant effect was observed for Experiment ($\beta = 0.07$, $t = 2.03$, $p = .043$), reflecting the generally longer turn durations in Experiment 1 (live confederate) than in Experiment 2 (pre-recorded confederate). Finally, a significant interaction was found between Task and Experiment ($\beta = –0.03$, $t = –2.18$, $p = .029$), reflecting the fact that the difference between the tasks was larger in Experiment 1 than in Experiment 2.

Tapping performance

The participants were asked to continuously tap throughout the Tapping and Speaking test block. Their tapping performance during the linguistic task was compared with their baseline performance during a 2-s time window immediately preceding the onset of the confederate's turn. Because of the high sensitivity of the microphone buttons, sometimes button presses as well as button releases activated the button-response trigger. Therefore, all instances where the same button-trigger was activated twice within 400 ms were discarded (25.0% and 27.5% of the button-triggers in Experiment 1 and 2, respectively). Furthermore,

5.1% of the trials in Experiment 1 and 3.8% of the trials in Experiment 2 were excluded from the analyses because the average tapping rate during the baseline period fell below one tap per second, which indicated a failure to perform the tapping task appropriately on that trial. For the remaining data, a button press was labelled as correct if it followed the correct predecessor. For instance, for the correct pattern 1−3–2−4, the predecessor for button 3 had to be 1, and for button 2, it had to be 3. The first tap in a block was always coded as correct. In Experiment 1, 15.5% of the valid taps, and in Experiment 2, 11.9% of the valid taps were discarded as incorrect. Tapping rate was calculated as the number of correct taps per second. Thus, it was an aggregated score, combining tapping speed and accuracy. Only trials featuring correct utterances (including non-fluent but ultimately correct trials) were included in the statistical analysis of the tapping performance. Based on this criterion, 5.3% of correct taps were discarded in Experiment 1 and 3.8% in Experiment 2.

Figure 3 shows the participants' average tapping performance across the trials. Data are time-locked to the end of the confederate's turn (time 0). The dotted line originating from the box labelled "base rate" indicates the average tapping rate during the 2-s interval immediately prior to the onset of the display. Note that for the calculation of the base rate, the base-window was aligned to the trial onset, not to speech offset as was the rest of the data in Figure 3. As can be seen, at the beginning of the confederate's turn, the tapping rate was close to the baseline in both experiments, but it deteriorated across the turn and was much lower during the participant's own turn.

For the statistical analyses, we first compared the tapping rates for two time windows (before and after the end of the confederate's turn) to the baseline. This analysis tested the prediction that tapping performance should be better at baseline (i.e., before confederate onset) than during listening or speaking. This analysis included main effects for Experiment (contrast coded) and Time Window, the latter being a categorical predictor with the levels Baseline (a 2-s window immediately preceding display onset; modelled on the intercept), Confederate Turn (from visual display onset to confederate offset), and Participant Turn (confederate offset to participants offset). The interaction between these fixed effects was also included. Adding Task Order did not improve the quality of fit. A random effects structure was included, involving intercepts for Participants and intercepts for Displays.

Confirming the visual impression, the optimal model ($\beta_{Intercept} = 3.01, t = 39.25, p < .001$) included a main effect of Time Window. Tapping rates differed from the baseline both during the confederate's turn ($\beta = -0.12, t = -5.61, p < .001$) and during the participant's turn ($\beta = -0.53, t = -23.49, p < .001$). The negative β estimates for these two windows demonstrate that tapping rates

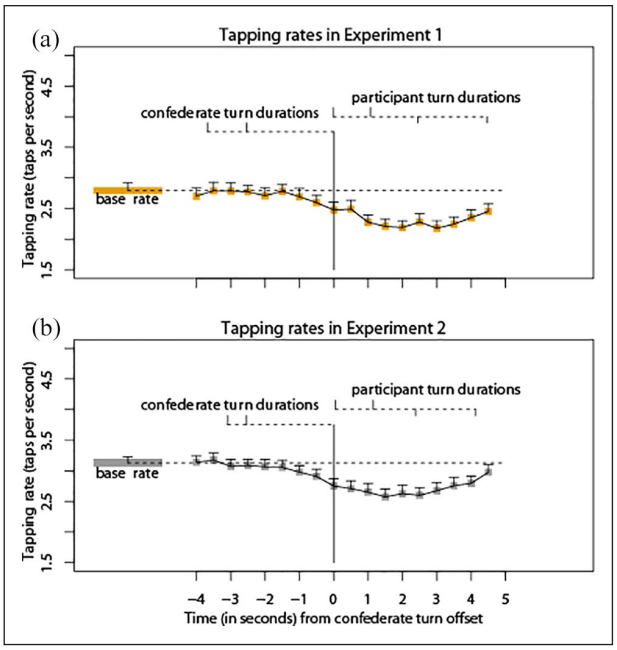

Figure 3. Average tapping rates across trials in (a) Experiment 1 (live confederate) and (b) Experiment 2 (pre-recorded confederate), aligned to the confederate's turn offset.
Error bars reflect standard errors of participant means. At the top of the panel, markers indicate turn durations, with upwards facing vertical markers indicating the 95% onset ranges (for both confederate and participant turns) and downwards facing markers indicating the 95% offset ranges for participant turns.

somewhat decreased, relative to the baseline, during the confederate's turn and especially deteriorated during the participant's own turn. The model also revealed a main effect of Experiment ($\beta = -0.18, t = -2.33, p = .02$), indicating that participants generally tapped more slowly in the presence of a confederate (Experiment 1) than when they heard a recording of her utterances (Experiment 2). There was no significant interaction between Experiment and Time Window.

To determine when participants started preparing their turns, a more detailed analysis was undertaken by dividing the time period around the confederate's offset (time 0) in 0.5-s steps. The model included the main effects Experiment (contrast coded) and Time Step (with levels Baseline [intercept] vs 18 separate steps, spanning across the trial duration from 4 s before confederate offset to 4.5 s after confederate offset). The random effects structure included intercepts for Participants and Displays. The model ($\beta_{Intercept} = 2.96, t = 35.21, p < .001$) showed that Bonferroni-corrected tests (corrected alpha = 0.0028) for the different levels of the factor Time Step (i.e., the different time steps compared with baseline) were significant at all time points from −0.5 to 4.5 s from confederate offset (all at $\beta_{Time[-0.5\ to\ +4.5]} \leq -0.21, t \leq -4.80, p < .001$). Uncorrected tests were significant from of −1 s from confederate offset onwards. Hence, participants experienced significant interference of speech planning on

their tapping performance from just before the offset of the preceding turn. A main effect of Experiment ($\beta = 0.17$, $t = 2.31$, $p < .02$) reflected lower tapping rates in the presence than in the absence of a confederate. The effect of time window was similar in both experiments, as the model showed no significant interactions between Experiment and any of the Time Steps. These results do not indicate any change in the timing of the participants' utterance planning in the presence or absence of the confederate.

Eye movement analyses

The participants' eye movements were analysed to determine how long they would preferentially attend to the objects mentioned by the confederate and when they would turn to the objects they had to describe themselves. We defined regions of interest (ROIs) for each of the 12 objects (eight line drawings and four arrows: ROIs included 0.5 cm around edges of the objects) and categorised the participants' fixations as falling onto any of the objects mentioned by the confederate (top row), onto any of the objects mentioned by the participant (bottom row), or elsewhere. Fixations with durations below 80 ms were discarded as spurious (2.0% of fixations in Experiment 1 and 1.4% of fixations in Experiment 2). While the objects were in view, the majority of the fixations fell within the pre-defined interest areas, either on one of the confederate's or one of the participant's objects (87.7% in Experiment 1 and 89.6% in Experiment 2). Thus, the fixation patterns for the two sets of objects were largely complementary: More gazes to the confederate's objects were accompanied by fewer gazes to the participant's objects.

Figure 4 displays the fixation preferences across the trials for each of the two sets of objects. A preference of zero means that participant and confederate objects were equally likely to be fixated. A preference of +1 means that only confederate objects were fixated, and a preference of of −1 means that only participant objects were fixated. Intermediate values indicate more or less pronounced preferences for confederate objects (positive values) or for participant objects (negative values). For this visualisation, samples during which participants fixated on neither the participants' nor the confederate's objects were coded as 0 (before the onset of the display, the preference is thus 0; see, for instance, time points before −4 s when no objects were in view). On average, the pictures appeared on the screen approximately 3.5 s before confederate offset. The confederate's turn began around time point −3 s, and the participants' turn ended around point +4 s.

As can be observed, during the confederate's turn, the participants' gaze depended highly on the task: in the Speaking-Only condition, there was a preference for the confederate's objects until about 1 s before the end of the confederate's turn. This preference appears to be somewhat more pronounced in the absence than in the presence

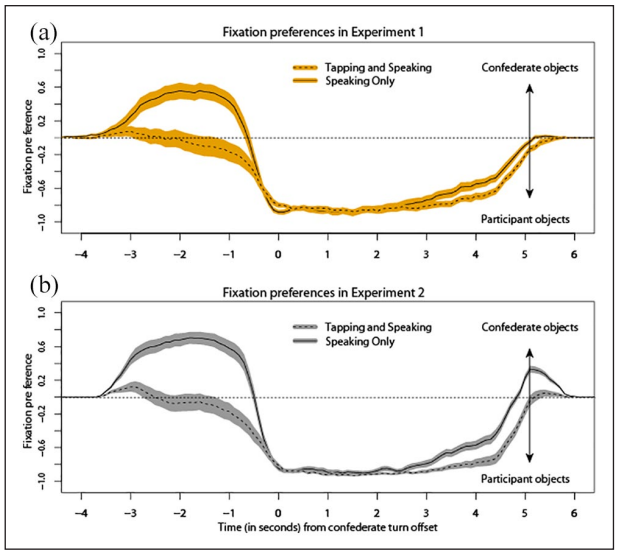

Figure 4. Fixation preferences for the objects described by the confederate (top section of each panel) or the objects described by the participant (bottom section of each panel) in the Tapping and Speaking and Speaking-Only condition of (a) Experiment 1 (live confederate) and (b) Experiment 2 (pre-recorded confederate).
Fixation preferences are aligned to the confederate's turn offset. Shaded error regions indicate the by-participant standard error of the mean.

of the confederate. By contrast, in the Tapping and Speaking condition, the participants never displayed a strong preference for the confederate's object but showed a preference for their own objects, which increased across the confederate's turn.

A first analysis of the gaze patterns compared the fixation proportions for two windows, before and after the end of the confederate's turn, respectively. This analysis tested the prediction that the proportion of fixations on the two rows of objects should change from the first to the second turn. The dependent variable was the proportion of fixations to confederate objects out of all fixations to the interest areas defined for both rows of objects. Main effects were Experiment (contrast coded), Time Window (contrast coded: Confederate Turn [from display onset to confederate offset] = 1; Participant Turn [from confederate offset to participant offset] = of −1), and Task (contrast coded). The interactions between these fixed effects were also included. Adding Task Order as a covariate did not improve the model. A random effects structure was included, involving intercept and slopes for Task for Participants and intercept for Displays.

Confirming the visual impression, the optimal model ($\beta_{Intercept} = 0.28$, $t = 18.16$, $p < .001$) included a main effect for Task ($\beta = 0.08$, $t = 9.99$, $p < .001$): fewer looks were directed at the confederate's objects in the Tapping and Speaking than in the Speaking-Only condition. There was also a main effect of Time Window ($\beta = 0.26$, $t = 125.11$,

$p < .001$): the number of looks to the confederate's objects decreased from the confederate's to the participant's turn. There was no main effect of Experiment ($\beta = -0.01$, $t = -0.88$, $p = .38$): the proportion of looks towards the confederate's items was very similar in the two experiments and thus not influenced by the presence or absence of the confederate.

In addition to the main effects of Task and Time Window, there were a number of interactions. Task interacted with Time Window ($\beta = 0.07$, $t = 33.59$, $p < .001$): task had a stronger effect on the participants' looks during the confederate's turn than during the participants' own turn, where the participants almost always fixated upon their own objects. A second interaction involved the factors Experiment and Time window ($\beta = -0.01$, $t = -6.19$, $p < .001$): in the first time window, when the confederate's objects were named, the participants looked slightly less often at these objects when the confederate was naming them live (Experiment 1) compared with when they were named in a recording (Experiment 2); in the second time window, that is, during the participants' turn, the participants rarely looked at the confederate's objects in either experiment.

There was no interaction between Experiment and Task ($\beta = -0.01$, $t = $ of -1.27, $p = .20$). However, there was a three-way interaction between Task, Experiment, and Time Window ($\beta = -0.02$, $t = -8.18$, $p < .001$). This indicates that the interaction of Experiment and Time window was weaker for the Tapping and Speaking than for the Speaking-Only condition.

In sum, these tests largely confirm the visual impression from Figure 4. In the absence of a secondary task, participants mostly fixated the confederate's objects during the confederate's turn, and then, during their own turn, predominantly fixated their own objects. In addition, participants were slightly less likely to fixate the confederate's objects during the confederate's turn when they were taking turns with the confederate than when they were listening to a recording of her utterances.

To determine more precisely when participants began to turn their attention from the confederate's to their own objects, we fitted separate a logistic regression model to the gazes of individual participants in each of the two tasks (Speaking-Only vs Tapping and Speaking). These models were fitted including a window from 2 s before to 2 s after the offset of the confederate's turn. The dependent variable was fixation preference (confederate's object coded as 1). From each of these models, the regression estimates were extracted and used to estimate the average 50% crossover point for each participant in each of the two tasks. In other words, we estimated when each participant began to show a stable preference for their own, rather than the confederate's objects. Since the crossover points were derived from regression estimates, they could be based on extrapolation beyond the 2-s window. Figure 5 displays the occurrence of the estimated crossover points for 0.5-s time bins around

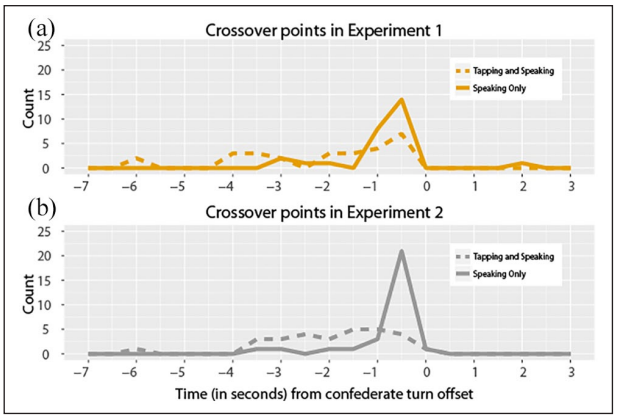

Figure 5. Distribution of crossover points (shift from fixation preference for confederate's to participants' own objects in (a) Experiment 1 (live confederate) and (b) Experiment 2 (pre-recorded confederate).
Time point 0 is the offset of the confederate's turn.

the offset of the confederate's turn (for the analyses, estimates were discretised as 0.1-s bins). As can be seen, in the Speaking-Only conditions of both experiments, most crossover points fell into a small time window starting around 1 s before the offset of the confederate's turn. By contrast, in the Tapping and Speaking conditions, the distribution of crossover points was much broader, and in Experiment 2, it did not show a clear peak at all.

A regression model was fitted to these estimates to statistically compare the crossover points across the conditions. This model included Experiment (contrast coded: Exp1 = 1; Exp 2 = of -1) and Task (contrast coded: SO = 1; TS = of -1) as predictors and Crossover Point as the dependent variable. The model ($\beta_{Intercept} = $ of -1.44, $t = $ of -12.75, $p < .001$) showed a significant main effect for Task ($\beta = 0.62$, $t = 5.53$, $p < .001$). The positive β reflects the fact that the crossover point occurred, on average, later in the Speaking-Only condition than in the Tapping and Speaking condition. No significant effect was observed for the factor Experiment or its interaction with Task. This indicates that the timing of the crossover points was independent of whether the confederate was present or absent.

Discussion

Experimental work in psycholinguistics has often used laboratory settings where participants carry out speaking or listening tasks in monologue settings, for instance, naming pictures without a listener or categorising recordings of spoken words or sentences in specific ways. A concern often raised against such studies is that speaking and listening most commonly occur in conversational settings and that studies not creating such settings may generate results that have little relevance for speaking and listening in everyday contexts. This concern cannot be addressed in general terms, since the mere presence of another person

or the interaction with them may have a stronger impact on performance in some tasks than in others. In this study, we returned to a paradigm used in an earlier study (Sjerps & Meyer, 2015) and asked how the speakers' speech planning strategies would differ when they responded to a confederate, who sat next to them in lab, or to pre-recorded utterances from the same speaker. In the earlier study, participants responded to recorded utterances. The results indicated that the speakers began to plan their utterances during the preceding turn, but not as early as predicted by Levinson and Torreira's (2015) Early Planning hypothesis. In discussions of these results, it has been suggested that speakers may plan their utterances earlier when they respond to another person rather than a recording (Barthel et al., 2017; Barthel et al., 2016; Bögels et al., 2018; Bögels, Magyari, & Levinson, 2015). This hypothesis was assessed in this study. The results demonstrated that the participants' planning strategies were strongly affected by whether or not they had to carry out a tapping task while preparing their utterances. In contrast, only modest effects of the presence or absence of the confederate were seen. In the remainder of this discussion, we first discuss the participants' speech output and then their tapping performance and their eye gaze pattern before turning to the broader implications of the findings.

Characteristics of participants' speech

The analyses of the participants' utterances, that is, the rates of hesitations, utterance onset latencies, and durations all showed significant interference from the concurrent tapping task. These results are not surprising. Speech planning requires cognitive capacity and is therefore hampered when combined with a secondary task, such as tapping or driving, which also requires cognitive capacity (e.g., Becic et al., 2010; Boiteau et al., 2014; Kubose et al., 2006). In addition, both tapping and speaking require the selection and initiation of motor responses, and these processes may have interfered with each other (e.g., Hegarty et al., 2000; Serrien, 2009). Finally, participants may have kept the instructions for the tapping task active in verbal working memory, which would have interfered with the retrieval of lexical items and their combination into phrases and with self-monitoring processes of the speech output (Klaus, Mädebach, Oppermann, & Jescheniak, 2017; MacDonald, 2016; Martin & Schnur, 2019; see also Larsen & Baddeley, 2003). An interesting question for further work concerns the precise origin of the performance decrement occurring during tapping. This line of work could have important implications for the choice of appropriate task combinations in further dual-task studies into speaking and listening.

In contrast to the marked effects of Task, Experiment (i.e., confederate presence or absence) only had a modest effect on the participant's speech output: there was no effect on the speech onset latencies or the fluency of the utterances. In other words, the prediction that speakers might begin to speak earlier in the presence than in the absence of the confederate was not borne out. There was an effect of Experiment on the durations of the utterances: in Experiment 1, when the confederate was present, the durations were longer and more strongly affected by the tapping task than in Experiment 2, where recordings were used. This interaction suggests that the presence of the confederate in some way increased the cognitive load of the task.

Tapping performance

The participants' performance in the tapping task was poorer in the presence than in the absence of the confederate. This supports the suggestion that confederate presence increased the cognitive load. How this increase in cognitive load arose is unclear. One possibility is that it was caused by the awareness of the presence of another person in the lab; participants may, for instance, have simulated their behaviour, which has been shown to cause interference (for discussion, see Gambi, Cop, & Pickering, 2015; Gambi, van de Cavey, & Pickering, 2015; Hoedemaker & Meyer, 2019). Another option is that the increase in cognitive load was related to properties of the utterances the participants heard. As described in the *Auditory Stimuli* section, the utterances presented in Experiments 1 and 2 stemmed from the same speaker and were well-matched for onset latencies and durations. The error rate in the confederate's speech was very low (no errors in the recordings, below 1% errors in the live descriptions), but there were some hesitations (on 10% of the trials) in the live descriptions but none in the recordings. This may have made listening to the descriptions slightly harder when the confederate was present than when they were absent. Note, however, that the tapping rate was lower in Experiment 1 than in Experiment 2 during both the confederate's and the participants' own turns. This suggests that the decrement was due to confederate presence per se, rather than properties of the speech input.

Importantly, the development of the tapping performance within the trials was not affected by confederate presence. Regardless of whether or not the confederate was present, the tapping rate was close to baseline at the beginning of the confederate's turn, then gradually decreased, and was significantly below baseline from about 1 s before the end of confederate's turn. This pattern replicates our earlier findings (Sjerps & Meyer, 2015) and indicates, first, that speech planning and speaking required more processing capacity and therefore interfered more with tapping than listening (see also Becic et al., 2010; Boiteau et al., 2014; Kubose et al., 2006) and, second, that speech planning began during the preceding turn. The gradual decrease in tapping scores

indicates that over the course of the confederate's turn, the participants became more and more likely to begin to plan their own utterance.

Eye movements

The analysis of the participants' eye movements showed that during their own turn, they almost exclusively looked at their own objects, regardless of whether or not they were tapping and regardless of the presence or absence of the confederate. This finding adds to a large body of findings from multiple-object naming studies demonstrating that speakers typically look at the objects they name or describe, most likely because this facilitates the recognition of the objects and the retrieval of their names (e.g., Hintz & Meyer, 2015; Meyer et al., 1998; Schotter, Ferreira, & Rayner, 2013; see also Coco & Keller, 2015; Henderson, Hayes, Rehrig, & Ferreira, 2018).

During the confederate's turn, the participants' gaze pattern depended strongly on the task, but very little on the presence or absence of the confederate. When the task was Speaking Only, the participants initially showed a strong preference for the confederate's objects and about a second before the end of the confederate's turn shifted their attention to their own objects. The preferences for the two sets of objects are shown in Figure 4. Figure 5 shows that in the Speaking-Only condition, the crossover points, that is, the time points when participants began to show a consistent preference for their own objects, fell into a narrow time window close to the end of the confederate's turn. Given the size of the objects and their spacing, the participants may have gleaned some information about the objects in the bottom row, their own objects, while fixating the confederate's objects in the top row. Yet, many studies have shown that speakers usually fixate upon the objects they have to name, even when they know the objects and can identify them extrafoveally (e.g., Malpass & Meyer, 2010; Meyer et al., 2012). As explained in the Introduction, directing one's visual attention at the relevant objects likely facilitates their recognition and the retrieval of the associated lexical information. Consequently, the viewing pattern in the Speaking-Only condition indicates that the participants initially focused their visual attention on the confederate's object and then, shortly before the end of the confederate's utterance, began to focus on their objects, and, most likely, began to plan their utterance about them. Confederate presence modulated the strength of the preference for confederate's objects: participants were slightly less likely to look at the confederate's objects and, conversely, more likely to look at their own objects, when the confederate was present in the lab than when they heard a recording. Nevertheless, during the confederate's turn, participants were far more likely to look at the confederate's than their own objects.

The gaze pattern in the Tapping and Speaking condition was different. The participants never showed a strong preference for the confederate's objects. Instead, they were initially about equally likely to look at the confederate's and their own objects and then, during the confederate's turn became more and more likely to look at their own objects. This can be seen in Figure 4 and in the absence of a clear peak in the crossover points in Figure 5. In contrast to the Speaking-Only condition, where participants quite uniformly turned to their own objects towards the end of the confederate's turn, the participants did not adopt a uniform processing strategy when they had to combine speaking and tapping. The gaze pattern does not reveal how the participants processed their own objects during the confederate's trial. For instance, participants may sometimes have briefly looked at one or several of their own objects and then returned their gaze to the confederate's objects, or they may have focused on one or two of their own objects and generated an utterance plan, to be launched after the end of the confederate's turn. The latter strategy would be most consistent with Levinson and Torreira's (2015) Early Planning hypothesis (see also Corps et al. (2018) for a relevant discussion of the distinction between utterance planning and launching).

In our earlier study (Sjerps & Meyer, 2015), we had observed a similar pattern, with speakers also being less likely to look at the objects mentioned in the recorded description in the Tapping and Speaking than in the Speaking-Only condition. However, the difference between the conditions was less pronounced than in this study. Most likely, this was because participants in the earlier study were encouraged more strongly to listen carefully to the descriptions they heard. This was because in addition to the conditions that were also included in this study, the earlier study included the Listening-Only condition, where participants listened to descriptions of both rows of objects but did not speak themselves. Moreover, in all conditions, they had to evaluate the correctness of the second turn of each trial, which they either heard or produced themselves. Although this judgement task did not apply to the first turn of the trial, the instructions may have encouraged the participants to pay close attention to all utterances and to look at most of the objects being mentioned. By contrast, in this study, paying close attention the confederate's utterances was not required. Taken together, the results of the two studies indicate that the cognitive load and the importance of paying attention to the confederate's utterances affected when the participants fixated upon the confederate's objects and when they turned to their own objects.

Differences in cognitive load and in the importance of listening carefully to the interlocutor undoubtedly also arise in the everyday conversations. Our results suggest that speakers respond to such demands by adjusting their processing priorities. In other words, it seems unlikely that there are default strategies for the allocation of attention to the interlocutor's or one own speech and for the coordination of

listening and speech planning; instead, there are probably many strategies speakers can employ as they see fit. Future work could aim to determine the variables that most strongly affect how people distribute their attention in linguistic dual-tasking, that is, when concurrent listening and speech planning are required.

Given that it has often been argued that early planning is crucial for smooth turn-taking in everyday conversation, one may ask how smooth turn-taking can be achieved if, as we argue, speakers do not necessarily plan their utterances early during their partner's turn. This is an important issue for further research, but we can offer some initial speculations. First, it is worth remembering that the gaps between turns are not uniformly short but vary considerably in duration. This has been shown for informal conversation (e.g., Heldner & Edlund, 2010). In other contexts (teaching, patient–doctor interactions, and scientific debates), there may be even more variability in gap durations. Thus, conversation might not always be as smooth as commonly assumed. Second, in conversations, speakers are not only free to decide when to begin to plan their utterances and when to launch them but also to determine what to say, both in terms of content and linguistic form. This is an important difference to laboratory situations where participants are asked to produce specific types of utterances at specific times. In everyday conversations, speakers must sometimes give precise answers to specific questions ("When is the next train to Amsterdam?"—"At 14:32"); but often a wide range of responses, from non-verbal and verbal back-channelling (head nod, "hm"), to questions for more information ("Really?"), comments ("How awful!"), and lengthy narratives ("The same happened to me when . . .") counts as relevant contributions. In addition, speakers can choose between multiple linguistic means (different referring expressions, syntactic structures) to express themselves. Such flexibility facilitates utterances formulation (e.g., Konopka & Meyer, 2014). Furthermore, utterance planning may be facilitated by various types of priming from the preceding context (e.g., Pickering & Garrod, 2004, 2013). Last but not least, speakers can plan their utterances incrementally and initiate them as soon as they have generated the first few words (e.g., Wheeldon & Lahiri, 1997).

Although the systematic changes of the planning strategy depending on the cognitive load and the importance of the listening task are novel and in our view interesting findings of this study, the main goal was to determine how the participants' planning strategies would be affected by the presence or absence of a confederate. We found subtle effects of confederate presence, but no evidence that participants adopted dramatically different planning strategies in the presence or absence of the confederate. These results were obtained in one specific paradigm. We do not suggest that experiments using speech recordings, enlisting the help of confederates, or testing pairs of naive participants will necessarily yield equivalent results. There is ample evidence to the contrary. For instance, Kuhlen and Brennan (2013) discuss how participants' behaviour may be affected by the presence of a confederate or a second naive participant, and several recent studies have demonstrated subtle but reliable effects of the presence or absence of a task partner on performance in speech production tasks (e.g., Gambi, Van de Cavey, & Pickering, 2018; Hoedemaker, Ernst, Meyer, & Belke, 2017; Hoedemaker & Meyer, 2019; Kuhlen & Abdel Rahman, 2017). An important task for future research is to determine exactly how social and linguistic variables jointly affect how language is comprehended and produced. As has been pointed out before (de Ruiter & Albert, 2017; Meyer et al., 2018), this can best be accomplished by combining observational data from analyses of spontaneous conversations with experimental work targeting specific aspects of comprehension or speech planning. Meanwhile, researchers planning studies of speaking or listening can only be advised to consider carefully how the presence or absence of an interlocutor might affect their participants' behaviour and choose their paradigm accordingly.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Antje S Meyer https://orcid.org/0000-0002-7735-9025

References

Almor, A. (2008). Why does language interfere with vision-based tasks? *Experimental Psychology*, *55*, 260–268. doi:10.1027/1618-3169.55.4.260

Barthel, M., Meyer, A. S., & Levinson, S. C. (2017). Next speakers plan their turn early and speak after turn-final "go-signals." *Frontiers in Psychology*, 8, Article 393. doi:10.3389/fpsyg.2017.00393

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, *7*, Article 1858. doi:10.3389/fpsyg.2016.01858

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes (R Package Version 1.1.10).

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin & Review*, *17*, 15–21. doi:10.3758/PBR.17.1.15

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Computer program). Available from http://www.praat.org

Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia*, *109*, 295–310. doi:10.1016/j.neuropsychologia.2017.12.028

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no. . . How the brain interprets the pregnant pause in conversation. *PLoS ONE*, *10*(12), e0145474. doi:10.1371/journal.pone.0145474

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*, 12881. doi:10.1038/srep12881

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57. doi:10.1016/j.wocn.2015.04.004

Boiteau, T. M., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology*, *143*, 295–311. doi:10.1037/a0031858

Cleland, A. A., Tamminen, J., Quinlan, P. T., & Gaskell, M. G. (2012). Spoken word processing creates a lexical bottleneck. *Language and Cognitive Processes*, *27*, 572–593. doi:10.1080/01690965.2011.564942

Coco, M. I., & Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *The Quarterly Journal of Experimental Psychology*, *68*, 46–74. doi:10.1080/17470218.2014.936475

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 886–899. doi:10.1037/0278-7393.34.4.886

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*, 84–107. doi:10.1016/0010-0285(74)90005-X

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, *175*, 77–95. doi:10.1016/j.cognition.2018.01.015

Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, *55*, 230–240. doi:10.1080/0163853X.2017.1330031

de Ruiter, J. P., & Albert, S. (2017). An appeal for a methodological fusion of conversation analysis and experimental psychology. *Research on Language and Social Interaction*, *50*, 90–107. doi:10.1080/08351813.2017.1262050

de Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*, 515–535. doi:10.1353/lan.2006.0130

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*, 210–233. doi:10.1016/0749-596X(91)90004-4

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187–1199. doi:10.1037/0278-7393.28.6.1187

Gambi, C., Cop, U., & Pickering, M. J. (2015). How do speakers coordinate? Evidence for prediction in a joint word-replacement task. *Cortex*, *68*, 111–128. doi:10.1016/j.cortex.2014.09.009

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1–21. doi:10.1037/a0037438

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

Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, *28*, 376–385. doi:10.3758/BF03198553

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversation. *Journal of Phonetics*, *38*, 555–568. doi:10.1016/j.wocn.2010.08.002

Henderson, J. M., Hayes, T. R., Rehrig, G., & Ferreira, F. (2018). Meaning guides attention during real-world scene description. *Scientific Reports*, *5*, 13504. doi:10.1038/s41598-018-31894-5

Hintz, F., & Meyer, A. S. (2015). Prediction and production of simple mathematical equations: Evidence from anticipatory eye movements. *PLoS ONE*, *10*(7), e0130766. doi:10.1371/journal.pone.0130766

Hoedemaker, R. S., Ernst, J., Meyer, A. S., & Belke, E. (2017). Language production in a shared task: Cumulative semantic interference from self- and other-produced context words. *Acta Psychologica*, *172*, 55–63. doi:10.1016/j.actpsy.2016.11.007

Hoedemaker, R. S., & Meyer, A. S. (2019). Planning and coordination of utterances in a joint naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 732–752. doi:10.1037/xlm0000603

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*, 151–171. doi:10.1016/j.actpsy.2010.11.003

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144. doi:10.1016/j.cognition.2002.06.001

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention and trans-saccadic memory. *Visual Cognition*, *5*, 127–155. doi:10.1080/713756783

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *The Quarterly Journal of Experimental Psychology*, *68*, 710–730. doi:10.1080/17470218.2014.964736

Kantola, L., & van Gompel, R. P. (2016). Is anaphoric reference cooperative? *The Quarterly Journal of Experimental Psychology*, *69*, 1109–1128. doi:10.1080/17470218.2015.1070184

Kemper, S., Herman, R. E., & Lian, C. H. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging*, *18*, 181–192. doi:10.1037/0882-7974.18.2.181

Kemper, S., Herman, R. E., & Nartowicz, J. (2005). Different effects of dual task demands on the speech of young and older adults. *Aging, Neuropsychology, and Cognition*, *12*, 340–358. doi:10.1080/138255890968466

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*, 255–289. doi:10.1080/0163853X.2014.955997

Klaus, J., Mädebach, A., Oppermann, F., & Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: Effects of concurrent verbal and visuospatial working memory load. *The Quarterly Journal of Experimental Psychology*, *70*, 811–831. doi:10.1080/17470218.2016.1167926

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*, 143–162. doi:10.1016/j.jml.2011.08.003

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

Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology*, *20*, 43–63. doi:10.1002/acp.1164

Kuhlen, A. K., & Abdel Rahman, R. (2017). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition*, *166*, 94–106. doi:10.1016/j.cognition.2017.05.024

Kuhlen, A. K., Bogler, C., Brennan, S. E., & Haynes, J. D. (2017). Brains in dialogue: Decoding neural preparation of speaking to a conversational partner. *Social Cognitive and Affective Neuroscience*, *12*, 871–880. doi:10.1093/scan/nsx018

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*, 54–72. doi:10.3758/s13423-012-0341-8

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin & Review*, *15*, 1135–1140. doi:10.3758/PBR.15.6.1135

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *The Quarterly Journal of Experimental Psychology*, *56*, 1249–1268. doi:10.1080/02724980244000765

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in Cognitive Sciences*, *20*, 6–14. doi:10.1016/j.tics.2015.10.010

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, Article 731. doi:10.3389/fpsyg.2015.00731

MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, *25*, 47–53. doi:10.1177/0963721415620776

Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*, 2530–2539. doi:10.1162/jocn_a_00673

Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, *8*, Article 211. doi:10.3389/fpsyg.2017.00211

Malpass, D., & Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 523–537. doi:10.1037/a0018522

Martin, R. C., & Schnur, T. T. (2019). Independent contributions of semantic and phonological working memory to spontaneous speech in acute stroke. *Cortex*, *112*, 58–68. doi:10.1016/j.cortex.2018.11.017

Mattys, S., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*, 203–243. doi:10.1016/j.cogpsych.2009.04.001

Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology*, *9*, Article 525. doi:10.3389/fpsyg.2018.00525

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33. doi:10.1016/S0010-0277(98)00009-2

Meyer, A. S., & Van der Meulen, F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*, 314–319.

Meyer, A. S., Wheeldon, L. R., Van der Meulen, F., & Konopka, A. E. (2012). Effects of speech rate and practice on the allocation of visual attention in multiple object naming. *Frontiers in Psychology*, 3, Article 39. doi:10.3389/fpsyg.2012.00039

Murfitt, T., & McAllister, J. (2001). The effect of production variables in monolog and dialog on comprehension by novel listeners. *Language and Speech*, *44*, 325–350. doi:10.1177/00238309010440030201

Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, *64*, 404–423. doi:10.1016/j.jml.2011.01.004

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

Pickering, M. J., & Garrod, S. (2013). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, *36*, 377–392. doi:10.1017/S0140525X12003238

R Core Team. (2013). R: A language and environment for statistical computing (Version 3.1.1). Vienna, Austria: R Foundation for Statistical Computing.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology*, *9*, 119–137. doi:10.1037/1076-898X.9.2.119

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, *2*, Article 307. doi:10.3389/fpsyg.2011.00307

Rommers, J., Meyer, A. S., & Praamstra, P. (2017). Lateralized electrical brain activity reveals covert attention allocation during speaking. *Neuropsychologia*, *95*, 101–110. doi:10.1016/j.neuropsychologia.2016.12.013

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735. doi:10.2307/412243

Sassa, Y., Sugiura, M., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2007). Cortical mechanism of communicative speech production. *NeuroImage*, *37*, 985–992. doi:10.1016/j.neuroimage.2007.05.059

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*, 393–414. doi:10.1017/S0140525X12000660

Schotter, E. R., Ferreira, V. S., & Rayner, K. (2013). Parallel object activation and attentional gating of information: Evidence from eye movements in the multiple object naming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 365–374. doi:10.1037/a0028646

Serrien, D. J. (2009). Verbal–manual interactions during dual task performance: An EEG study. *Neuropsychologia*, *47*, 139–144. doi:10.1016/j.neuropsychologia.2008.08.004

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*, 159–187. doi:10.1016/j.actpsy.2005.01.002

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. doi:10.1016/j.cognition.2014.10.008

Somberg, B. L., & Salthouse, T. A. (1982). Dividend attention abilities in young and old adults. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 651–663.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . .Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10587–10592. doi:10.1073/pnas.0903616106

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*, 462–466. doi:10.1111/1467-9280.00386

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, *50*, 23–51. doi:10.1080/0163853X.2012.727719

Tolins, J., Zeamer, C., & Fox Tree, J. E. (2018). Overhearing dialogues and monologues: How does entrainment lead to more comprehensible referring expressions? *Discourse Processes*, *55*, 545–565. doi:10.1080/0163853X.2017.1279516

Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: The time course of response planning in conversation. *Frontiers in Psychology*, *6*, Article 284. doi:10.3389/fpsyg.2015.00284

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356–381. doi:10.1006/jmla.1997.2517