

Planning and Coordination of Utterances in a Joint Naming Task

Renske S. Hoedemaker

Max Planck Institute for Psycholinguistics, Nijmegen,
the Netherlands

Antje S. Meyer

Max Planck Institute for Psycholinguistics, Nijmegen, the
Netherlands, and Radboud University Nijmegen

Dialogue requires speakers to coordinate. According to the model of dialogue as joint action, interlocutors achieve this coordination by corepresenting their own and each other's task share in a functionally equivalent manner. In two experiments, we investigated this corepresentation account using an interactive joint naming task in which pairs of participants took turns naming sets of objects on a shared display. Speaker A named the first, or the first and third object, and Speaker B named the second object. In control conditions, Speaker A named one, two, or all three objects and Speaker B remained silent. We recorded the timing of the speakers' utterances and Speaker A's eye movements. Interturn pause durations indicated that the speakers effectively coordinated their utterances in time. Speaker A's speech onset latencies depended on the number of objects they named, but were unaffected by Speaker B's naming task. This suggests speakers were not fully incorporating their partner's task into their own speech planning. Moreover, Speaker A's eye movements indicated that they were much less likely to attend to objects their partner named than to objects they named themselves. When speakers did inspect their partner's objects, viewing times were too short to suggest that speakers were retrieving these object names as if they were planning to name the objects themselves. These results indicate that speakers prioritized planning their own responses over attending to their interlocutor's task and suggest that effective coordination can be achieved without full corepresentation of the partner's task.

Keywords: dialogue, joint action, corepresentation, object naming, eye tracking

Dialogue is an impressive feat of collaborative improvisation. In a dialogue, interlocutors must continuously process new information and formulate appropriate responses, all while maintaining smooth turn transitions with short interturn pauses (De Ruiter, Mitterer, & Enfield, 2006; Stivers et al., 2009). In addition, dialogue is a highly collaborative activity, in which interlocutors jointly construct meaning by building upon each other's utterances (Brennan, Galati, & Kuhlen, 2010; Clark, 1996; Garrod & Anderson, 1987). The tight coordination of actions observed during dialogue suggests that the interlocutors' underlying cognitive processes are also closely coordinated. This means that dialogue requires speakers to manage the within-speaker processes involved in speech production in parallel with processes of between-speaker coordination (Brennan et al., 2010). For example, while listening to a conversation partner, a speaker may be planning what to say next and trying to predict the upcoming end of their partner's turn

(e.g., Bögels, Magyari, & Levinson, 2015; Levinson & Torreira, 2015). The mechanisms that allow for such coordination and the ways in which interlocutors manage both between-speaker coordination and within-speaker speech-planning processes during dialogue are topics of growing interest and debate.

Arguably, the only model specifying a mechanism by which interlocutors coordinate is the model of dialogue as joint action (Gambi & Pickering, 2011; Garrod & Pickering, 2009). Any situation in which two or more individuals coordinate their actions to achieve a shared goal may be considered a form of joint action, including many everyday activities such as playing doubles in tennis or moving a heavy couch (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Bekkering, & Knoblich, 2006; Wenke, Atmaca, Holländer, Liepelt, Baess, & Prinz, 2011). To coordinate their behavior, coactors must predict each other's upcoming moves and integrate this information into the planning and execution of their own actions (Clark, 1996; Sebanz & Knoblich, 2009; Wilson & Knoblich, 2005). Successful predictions of this kind rely on an individual's understanding of the shared goal and a representation of each individual's task share within the context of the task as a whole (see also Bekkering, De Bruijn, Cuijpers, Newman-Norlund, Van Schie, & Meulenbroek, 2009; Hommel, 2009). Studies on joint tasks have shown that coactors take aspects of their partner's task share into account even when this is not strictly required for successful performance of their own portion of the task (e.g., Atmaca, Sebanz, & Knoblich, 2011; Atmaca, Sebanz, Prinz, & Knoblich, 2008; Böckler, Knoblich, & Sebanz, 2012; Sebanz, Knoblich, & Prinz, 2003). This joint interpretation of a shared task has been proposed to reflect coactors representing at

This article was published Online First July 12, 2018.

Renske S. Hoedemaker, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; Antje S. Meyer, Max Planck Institute for Psycholinguistics, and Donders Centre for Brain, Cognition, and Behavior, Radboud University Nijmegen.

We thank Caitlin Decuyper, Annelies van Wijngaarden, Eirini Zormpa, and other members of the Psychology of Language Department at the Max Planck Institute for Psycholinguistics for their help with this project.

Correspondence concerning this article should be addressed to Antje S. Meyer, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, the Netherlands. E-mail: antje.meyer@mpi.nl

least some aspects of each other's task as if it were their own (Holländer, Jung, & Prinz, 2011; Prinz, 1997; Sebanz et al., 2006;).

The model of dialogue as joint action (Gambi & Pickering, 2011; Garrod & Pickering, 2009) applies this account of joint action to dialogue. According to the model, the close coordination of utterances during dialogue resembles the coordination of actions in nonlinguistic joint tasks, meaning it relies on a similar process of corepresentation. Indeed, a large body of literature suggests that language users predict upcoming linguistic input during reading and listening (e.g., Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kuperberg & Jaeger, 2016; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005), and several prominent accounts of predictive processing during language comprehension propose a close involvement of the production system (e.g., Dell & Chang, 2013; Federmeier, 2007; Pickering & Garrod, 2007, 2013). Specifically, the model of language as joint action proposes that comprehenders generate forward models of their interlocutor's upcoming utterances in the same way they generate projections for their own utterances during speech planning (Pickering & Garrod, 2007, 2013). As a result, upcoming utterances come to be represented by both interlocutors regardless of who will produce the utterance, resulting in a set of shared representations from which both coactors operate. This corepresentation is what allows for the fast-paced, smooth interactions that characterize conversation.

Several recent studies provide evidence of the production system's involvement in the anticipation of a coactor's utterances. In a joint object-naming study by Baus and colleagues (2014), participants took turns naming objects together with a partner. Event-related potentials (ERPs) showed an effect of object-name frequency both when the participant was preparing to name the object and when they were waiting for their partner to name the object, but not when both speakers were instructed to ignore the object. The fact that participants appeared to engage in a process of lexicalization while they were anticipating their partner's utterance, even though they were not preparing to speak themselves, suggests that similar mechanisms are involved in both the planned production and predicted perception of an upcoming object name. Along the same lines, Hoedemaker, Ernst, Meyer, and Belke (2017) showed that both naming an object and listening to a partner name an object had an inhibitory effect on the subsequent naming of objects from the same semantic category. This cumulative slowing of object naming latencies within categories, also known as cumulative semantic interference, is a robust effect in single-speaker object naming and is considered to reflect a form of implicit long-term learning (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010; see also Kuhlen & Abdel Rahman, 2017). The observation that such learning took place following perceived (i.e., interlocutor-produced) as well as self-produced utterances suggests that speaking and listening in this context engaged similar processing mechanisms. However, in both the Baus et al. (2014) and Hoedemaker et al. (2017) studies, individual trials served to delineate speaker turns. In other words, on each trial an individual participant's task was either to respond or to wait while their partner responded to the displayed stimulus, and the speakers never responded directly to each other. As a result, the task did not require any direct between-speaker coordination and the planning of a speaker's own responses never overlapped with the anticipation of their partner's response. There-

fore, the evidence that predictive processing of linguistic input can engage the production system in these tasks does not necessitate that such a process occurs in dialogue.

Gambi, Cop, and Pickering (2015a) have provided evidence that speakers can corepresent each other's task when they prepare to speak simultaneously. In a joint word-replacement task, speakers found it harder to inhibit the production of a previously planned word when they knew their partner was simultaneously preparing to speak. The single-speaker version of the word-replacement task is a picture-naming task in which on some trials the initial picture is quickly replaced by another picture (Hartsuiker, Catchpole, De Jong, & Pickering, 2008; Tydgate, Diependaele, Hartsuiker, & Pickering, 2012; Tydgate, Stevens, Hartsuiker, & Pickering, 2011). When this happens, participants are either instructed to stop naming the initial picture and instead name the replacement picture, or to simply stop naming the initial picture and ignore the replacement. In Gambi et al.'s (2015a) joint word replacement task, participants found it more difficult to stop naming the initial picture when their partner was supposed to name the replacement picture ("other" condition) compared with when nobody was to name the replacement picture ("no" condition), although not as difficult as when they had to name the replacement picture themselves ("self" condition). This pattern suggests that representing a partner's task to begin speaking interfered with the participant's own task to stop speaking. The observed interference may reflect a language-specific process by which simulation of the partner's speech planning process resulted in activation of the participant's own production system. Alternatively, the results may reflect a more general go/no-go conflict, as the representation of the partner's go signal interferes with the participant's own stop signal (but see Demiral, Gambi, Nieuwland, & Pickering, 2016). Finally, the unexpected replacement of the initial picture and resulting change in speaker may cause conflict related to agent identification, that is, determining whose turn it is to speak (see Wenke et al., 2011). Regardless of the specific mechanism responsible, based on the graded pattern of results (the "other" condition showed interference relative to the "no" condition but less so than the "self" condition) the authors proposed a partial corepresentation account, suggesting speakers use some, but not all, of the same mechanisms they use to prepare their own speech to represent their interlocutors' speech planning process.

Interference in joint language tasks has also been observed when the coactors performed the same rather than conflicting tasks. Gambi, Van de Cavey, and Pickering (2015b) showed that participants were slower to name pictures when they believed their task partner, who was seated in another room such that they could not hear each other or otherwise interact, was simultaneously preparing to name the same pictures. Interestingly, the amount of interference was the same regardless of whether the interlocutors were preparing the same or different utterances, suggesting the interference was a result of the awareness that another actor was preparing to speak rather than knowing what they were preparing to say. Nonetheless, no interference was observed when the task partner was semantically categorizing (indicating their answer with a verbal yes/no response) rather than naming the pictures (see also Richardson et al., 2012, for a demonstration of a related "same-task" effect in a joint perception paradigm). Therefore, it appears that the specific representation of a partner engaging in a process of lexicalization interfered with the speaker's own naming re-

sponse, even when the content of the individual task shares did not conflict because both partners were preparing to name the same objects.

In summary, the available evidence suggests that an individual speaker's performance may be affected by the awareness that an interlocutor is also engaged in speech planning. These findings support the notion that speakers corepresent each other's operations and that this involves some of the same resources as planning one's own speech. However, the studies investigating this phenomenon so far have used experimental paradigms that require interlocutors to take turns responding to a stimulus (e.g., a picture on the screen) rather than directly to each other (Baus et al., 2014; Gambi et al., 2015a; Hoedemaker et al., 2017) or asked participants to respond to stimuli simultaneously while they could not hear each other (Gambi et al., 2015b). In dialogue, speakers each perform complementary portions of the shared task and respond to each other in a consecutive, turn-taking fashion. In addition, the content and timing of consecutive utterances is highly contingent upon one another and interlocutors must be able to process their partner's current utterance and engage in the planning of their own upcoming utterance at the same time (Barthel, Sauppe, Levinson, & Meyer, 2016; Bögels et al., 2015; Levinson & Torreira, 2015). As a result, any anticipation or corepresentation of a partner's task must occur in parallel with the speaker's own speech planning (Brennan et al., 2010). Such parallel processing might entail that participants continuously divide their attention between preparing their own speech plan and corepresenting the partner's plan; it could involve rapid switches between the two processes; or, as proposed by the model of language as joint action (Gambi & Pickering, 2011; Garrod & Pickering, 2009), it could involve the generation of a single integrated set of representations that capture both interlocutors' contributions.

The current study investigates how speakers manage the parallel pressures of between-speaker coordination and within-speaker speech planning, and aims to contribute to a better understanding of what it means to corepresent another person's speech planning in a joint verbal task. We used a novel multiple-object naming task in which pairs of speakers had the shared goal of producing fluent multiword utterances (for instance *car, bike, drum*) with minimal pausing in between words. The paradigm builds upon the earlier joint naming studies but incorporates several aspects relevant to the process of between-speaker coordination that were not present in these studies: the two speakers were seated in the same room

and could see and hear each other; they took turns rather than respond simultaneously; and they were asked to coordinate responses directly with each other (i.e., timing their utterance to the interlocutor's turn end). Structurally, this task resembles question-answer sequences or collaborative turn-completions regularly observed in normal dialogue (Lerner, 1991). In such situations, speakers are able to coordinate their responses in time, as evidenced by the short interturn pauses typically observed in normal conversation (De Ruiter et al., 2006; Stivers et al., 2009).

Following previous experiments on joint action, we compared behavior in the joint naming conditions to conditions in which speakers named the objects alone (e.g., Atmaca et al., 2008, 2011; Böckler et al., 2012; Gambi & Pickering, 2011; Sebanz et al., 2003). The model of dialogue as joint action predicts that their shared goal to coordinate utterances will result in the interlocutors corepresenting each other's task share. A strong hypothesis derived from this model is that speakers should process the objects named by their partner in exactly the same way as "their own" objects, only inhibiting the overt articulation of the object names. On this view, a speaker should attend equally to objects that will be named by their partner as to objects they plan to name themselves. Moreover, their latencies to initiate an utterance should be equivalent for utterances in which they name two objects and for utterances in which they name the first of a pair of objects and the partner names the second object. An alternative hypothesis is that speakers might pay some attention to their partner's objects and utterances but strongly prioritize their own objects and utterances.

Experiment 1

Experiment 1 investigated whether or not speakers in an interactive joint object-naming task showed corepresenting behavior (see also Gambi & Pickering, 2011, who proposed a similar experiment). For each experimental session, two speakers (A and B) were invited to the lab to perform the object naming task together. On each trial, three objects were displayed side by side (see Figure 1). The experiment included four naming conditions. Two of these were individual naming conditions, in which only Speaker A spoke, whereas Speaker B observed silently: in the Axx condition, Speaker A named the first (leftmost) object and in the AAX condition, Speaker A named both the first and the second (middle) object. Studies with individual speakers have shown that planning to name two pictures takes longer than planning to name

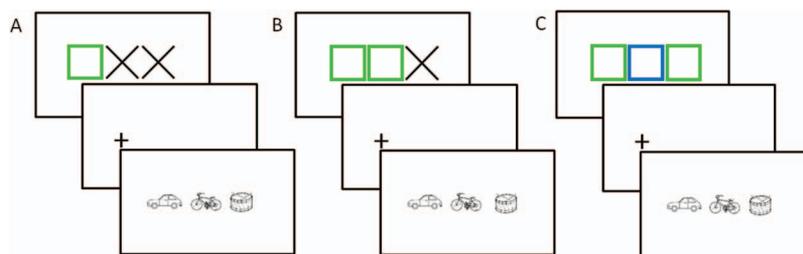


Figure 1. Example of the stimulus presentation in Experiment 1. The boxes preceding the pictures instructed Speaker A (green (shown as grey) boxes) and Speaker B (blue (shown as black) box) which objects they should name. Panels A to C show the instruction screens for the Condition Axx, AAX, and ABA, respectively. The ABX condition and filler conditions (Bxx, BBx, BAX, and BAB) are not depicted. See the online article for the color version of this figure.

one (Levelt & Meyer, 2000; Meyer, 1996; Meyer, Sleiderink, & Levelt, 1998) and that speakers typically move their eyes from the first to the second object before initiating the first of the two object names (Gordon & Hoedemaker, 2016; Griffin, 2001; Meyer, Belke, Häcker, & Mortensen, 2007; Roelofs, 2007, 2008). The interpretation of these findings is that in order to ensure a fluent production of the two names, speakers do not simply begin to speak when the first object name is available for pronunciation, but begin to plan the name of the second object before initiating their utterance. Behavior in the individual naming conditions of the current experiment was expected to match these established findings: we expected speech onset latencies to be longer in the AAX relative to the Axx condition and gaze shifts from Object 1 to Object 2 to occur before speech onset in the AAX but not the Axx condition.

The first joint naming condition was the ABx condition, in which Speaker A named the first object, followed by Speaker B naming the second object. If Speaker A corepresents B's task, this should result in their behavior reflecting some of the same operations as if they were themselves planning to name both the first and the second object (as they do in the AAX condition), resulting in longer latencies to speak and gaze shifts from the first to the second object occurring before speech onset. In contrast, if Speaker A attends only to their own planning requirements, we expect their behavior in the ABx condition to resemble that when they are only naming the first object and B is to remain silent (as in the Axx condition).

To assess whether corepresentation may be encouraged when it is potentially beneficial to individual task performance, we included a second joint naming condition. In the ABA condition, Speaker A named the first object, Speaker B named the second object, and Speaker A then named the third and final object. The previously discussed ABx condition imposes the shared task goal of producing a fluent utterance with minimal interword pausing. Therefore, the model of dialogue as joint action predicts that corepresentation should occur. However, this condition is not interactive in the sense that Speaker B's behavior does not affect any aspects of A's task, which may make A unlikely to take B's task share into account when planning their own responses. This result would be unexpected given earlier findings using shared tasks (e.g., Atmaca et al., 2008, 2011; Böckler et al., 2012; Sebanz et al., 2003) but not implausible, as corepresenting B's task would provide no obvious benefit to Speaker A in the ABx condition. In contrast, in the ABA condition corepresentation of Speaker B's speech planning operations may serve as a reasonable estimate for the timing of B's response, which could aid Speaker A with the precise timing of their final response following B. If this potential self-benefit encourages corepresentation, we expect to observe longer latencies to speak and earlier gaze shifts relative to speech onset in the ABA compared with the Axx and ABx conditions. However, A's task in the ABA condition is also more complex than in the AB condition (having to name two objects instead of one), which may encourage Speaker A to focus more exclusively on their own task and pay less attention to B's.

Finally, we manipulated the semantic relationship between the first and second object, such that on each trial they either belonged to the same or different semantic categories. Single-speaker naming studies have found that naming two related objects results in semantic interference compared with naming two unrelated objects

(Aristei, Zwitterlood, & Abdel Rahman, 2012; Freedman, Martin, & Biegler, 2004; Veenstra, Meyer, & Acheson, 2015). However, non-named distractor pictures typically do not interfere with the naming of semantically related targets (e.g., Damian & Bowers, 2003; Meyer & Damian, 2007; Navarrete & Costa, 2005; Roelofs, 2008), unless the distractor is sufficiently attended (Jescheniak, Matushanskaya, Mädebach, & Müller, 2014; Matushanskaya, Mädebach, Müller, & Jescheniak, 2016). We explored whether semantic interference in the form of longer naming latencies and viewing times for the first object in the related than in the unrelated condition could be found in the current paradigm and, if so, whether it would be equally pronounced in all naming conditions. Within the joint-action literature, it has not yet been established what exactly coactors represent about their partner's task. According to the task corepresentation account (Sebanz et al., 2006), coactors represent not only that their coactor is planning to act, but also the specific stimulus-response (S-R) mappings of their coactor's task share. In the case of joint action in language tasks, this has been interpreted to mean that interlocutors represent the lexical content of the utterance their partner is planning (Baus et al., 2014). Alternatively, according to the actor corepresentation account (Gambi et al., 2015b; Wenke et al., 2011), coactors represent the fact that their partner is acting or about to act, but not the S-R mappings or content of their response. We reasoned that if Speaker A represents the content of Speaker B's planned utterance on joint naming trials (task corepresentation), this should result in semantic interference even when A only names the first object, and the second object is part of B's task. In contrast, if Speaker A's representation of Speaker B includes their planning to speak but not the content of their utterance (actor corepresentation) we expect semantic interference to affect Speaker A's naming when they name both objects themselves but not in the joint-naming conditions.

Method

Participants. Thirty-four pairs of native speakers of Dutch participated in the experiment. One pair was excluded because of a technical problem, leaving a total of 33 pairs in the analysis. Twenty-two pairs consisted of two female participants and the remaining 11 pairs consisted of one male and one female participant. Average participant age was 23.2 years (range = 19–30 years). Ethical approval to conduct the study was given by the Ethics Board of the Social Sciences Faculty of Radboud University.

Materials and design. On every trial, the participant was presented with three black and white line drawings of objects placed side by side (see Figure 1). A total of 80 semantically related object pairs were selected for the experiment. The majority was taken from the Severens, Van Lommel, Ratinckx, and Hart-suiker (2005) picture naming database, with additional items taken from the picture database of the Max Planck Institute for Psycholinguistics. Semantic relatedness was defined as belonging to the same superordinate semantic category. Critical pairs were presented in position 1 (the first, left-most object of each triplet) and 2 (the second, middle object) on each trial. An additional 80 unrelated objects were selected from the same sources to appear in position 3 (the third, right-most object in each triplet). All experimental triplets are listed in the Appendix. Unrelated pairs were

generated by repairing related and unrelated objects, whereas the filler object in position 3 was kept constant. Table 1 presents the word frequency (SUBTLEX-NL corpus; Keuleers, Brysbaert, & New, 2010), age of acquisition (Severens et al., 2005; norms available for 183 items), and length in number of phonemes for the items in the three positions. In assigning objects to the first or second position, we aimed to place the object that was likely to be easier to name in the first position, as this should encourage participants to process the two object names in parallel (e.g., Malpass & Meyer, 2010). A single object naming experiment carried out with 24 participants recruited from the same pool as described above showed that the mean naming latencies were shorter ($B = 65.52$, $SE = 19.12$, $t = 3.43$) for the objects in first, compared to those in second position. All objects were adjusted to fit within a virtual square of 170×170 pixels (4° of visual angle) and were arranged such that the distance between the centers of two adjacent pictures was always 170 pixels or 4° of visual angle.

The object triplets appeared in four naming conditions. In the Axx condition, Speaker A was instructed to name only the first object of the triplet (e.g., *car*). In the AxA condition, Speaker A was instructed to name the first and second object, starting with the first object (e.g., *car bike*). In these individual naming conditions, Speaker B simply observed silently. In the ABx condition, Speaker A named the first object, followed by Speaker B naming the second object and nobody named the third object. In the ABA condition, Speaker A named the first object, Speaker B the second object, and Speaker A the third object. Relatedness of the first and second object and naming condition were counterbalanced across items in a 2 (related vs. unrelated) by 4 (Axx, AxA, ABx, ABA) within-participants design.

The experimental object pairs were distributed across four lists, such that each first object appeared twice per list; once followed by a related second object and once followed by an unrelated object. Within a list, both occurrences of an object appeared in the same naming condition, whereas across lists, each object appeared in each of the four critical naming conditions. Finally, 80 filler trials in which Speaker B named the first object (naming Conditions

Bxx, BBx, BAx, and BAB) were added to each list. Per list, every object appeared once in a filler trial, either in the related or in the unrelated condition, and in one of the four possible filler conditions. In total, each list consisted of 240 naming trials, with an equal number of related and unrelated trials and an equal number of trials in each naming condition. By an automatic procedure in the EyeLink software all items in a list appeared in a different random order for each pair of participants. As noted, objects always appeared in the same position, meaning for example that a position 1 object would never appear in position 2 or 3. Eight practice trials were added to the start of each list. Objects on the practice trials did not occur elsewhere on the list and data from the practice trials was not included in the analyses.

Apparatus and procedure. The stimuli were presented on two 24-in. display monitors (1920×1080 resolution) positioned side by side. Participants were seated each in front of their own monitor, approximately 67 cm from the display. They could see and hear each other and were aware that their monitors always displayed the same information. At the beginning of the experimental session, one participant was randomly assigned the role of Speaker A. Speaker A was seated at the left display and during the experiment his or her eye movements were recorded from the dominant eye using an EyeLink 1000 (SR Research, Ottawa, Ontario (CA)) table-mounted eye tracker. The other participant in the pair was assigned the role of Speaker B and seated at the display to the right of Speaker A. Speaker B's eye movements were not recorded. Vocal responses were recorded from both participants using a microphone and ASIO soundcard.

Each experimental session began with a familiarization phase, in which the participants were presented with all experimental objects in a pdf document, each with its name printed underneath. Participants could study the objects individually at their own pace. The test phase started with a 9-point calibration procedure for Speaker A. After the initial calibration, each trial began with a drift check, and calibration was redone if necessary. Figure 1 illustrates the sequence of events on each trial. The start of the trial was marked by a fixation point (drift check). Once this point was fixated, an instruction screen was presented for 2000 ms, indicating which objects should be named and by which speaker on that trial. On the instruction screen, green squares were used to indicate objects to be named by Speaker A and blue squares to indicate objects to be named by Speaker B. A black X was used to indicate that an object in that position should not be named. Subsequently the participants saw a fixation cross presented for 1,000 ms in the location of position 1, followed by a 500-ms blank screen, followed by the naming display containing the to-be-named objects. Participants were instructed to name the object(s) as quickly and accurately as possible without pausing between consecutive names. The naming display timed out after four seconds in conditions in which one or two objects were named and after six seconds in conditions in which three objects were named. Each experimental session took approximately 75 min in total.

Analysis. Most of the results reported below are based on measurements from Speaker A, whose eye movements were recorded and who named the object in position 1 on the 160 critical trials. Speaker B only named the object in position 1 on 80 trials, and semantic relatedness between the first and second object name was counterbalanced across participants rather than being tested within items and within participants. For these reasons, the trials

Table 1
Experiments 1 and 2: Average Log Frequency (SUBTLEX-NL, Keuleers et al., 2010), Age of Acquisition (AoA), Length in Phonemes, Single Object Naming Latency (RT), and Error Rate for Objects in Positions 1, 2, and 3

Item properties	Position 1 <i>M</i> (<i>SD</i>)	Position 2 <i>M</i> (<i>SD</i>)	Position 3 <i>M</i> (<i>SD</i>)
	Experiment 1		
Log frequency	3.13 (.67)	2.48 (.57)	2.79 (.74)
AoA	5.27 (.96)	5.69 (1.08)	5.67 (1.25)
Phonemes	4.09 (1.37)	4.80 (1.66)	4.57 (1.72)
RT	792ms (193)	850ms (230)	841ms (224)
Error rate	7%	12%	7%
	Experiment 2		
Log frequency	3.08 (.66)	2.48 (.58)	2.79 (.74)
AoA	5.32 (1.01)	5.68 (1.08)	5.67 (1.25)
Phonemes	4.05 (1.44)	4.80 (1.66)	4.57 (1.72)
RT	787ms (192)	842ms (221)	841ms (224)
Error rate	5%	10%	7%

where Speaker B was the first speaker were regarded as filler trials. Nevertheless, we also report analyses of Speaker B's naming latencies.

Two items were removed because of a stimulus error. Fixations shorter than 80 ms and within one degree of a longer, immediately subsequent fixation were merged with the longer fixation by an automatic procedure in the SR Dataviewer software. Areas of interest were drawn around each object according to a virtual 170×170 pixel box, such that the boundary between objects occurred at 85 pixels (2° of visual angle) from the center of each object.

First-pass viewing time for each object was measured as the time spent fixating each object before moving the eyes to a new object, either to the right or left (equivalent to what is typically termed *gaze duration* in studies of reading). *Eye-voice lag* was determined as the time between the offset of the last fixation on position 1 and the onset of the vocal response naming the object in position 1. The onsets and offsets of all vocal responses were manually determined by trained coders using Praat (Boersma & Weenink, 2017). *Naming latency* was measured from the onset of the object display to the onset of the first object name. *Pause time* was measured as the time between the vocal offset of one object name and the vocal onset of the next object name.

Trials on which either speaker made a naming error (6.2% of the trials, see also below) were excluded from the analysis. Naming errors were defined as any trials where at least one of the following events occurred: one or more objects were named incorrectly (63% of the error trials) or omitted (12% of the error trials; the omitted name was usually the last name in the utterance), one of the speakers hesitated or corrected the utterance (42% of the error trials), or speakers failed to follow the instructions (4% of the error trials) by both naming the same object, naming an object that should not have been named, or initiating their turn so early that the partner responded to the interruption, usually with laughter.

Even though the main analyses only report measurements from Speaker A, we also excluded trials on which Speaker B made an error because an erroneous response by Speaker B might affect Speaker A's eye movements (e.g., [re]fixating objects upon hearing an unexpected response from Speaker B) or vocal responses (e.g., cases in which B named an object that was meant to be named by A). On 6.7% of correct critical trials, Speaker A was not fixating the object in position 1 at the onset of the trial. These trials were removed from all subsequent analyses. In addition, trials on which the latency to begin speaking or the first-pass viewing time for the first object differed more than 2.5 standard deviations from the mean in the relevant naming and relatedness condition were removed from the analysis, as were trials on which the first-pass viewing time on the first object was less than 100 ms, resulting in a total of 5.2% of remaining trials being trimmed from the dataset.

The data were analyzed with linear mixed-effects models implemented using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R (R Core Team, 2014). Unless otherwise mentioned, all models were fit on log transformed RTs and the random effect structure included random intercepts for participants and items, and random slopes for the effects of naming condition and relatedness by participants and by items. Naming condition was included as a fixed effect categorical predictor with four levels, dummy coded with Axx as the initial reference condition and relevelled whenever necessary to make additional comparisons.

Relatedness was included as a fixed effect categorical predictor with two levels, sum-to-zero coded. Analyses of binary outcome variables (e.g., fixation probability) were done using logistic regression models that used the BOBYQA optimizer (Powell, 2009).

Results

Naming accuracy. Accuracy rates were computed as the proportion of trials that contained no naming errors from either speaker. Accuracy was generally high indicating that speakers successfully performed the task. Average naming accuracy across participant pairs was 93.8% (range = 86.5%–100%). Unsurprisingly, trials with more words uttered during the trial contained more errors. Accuracy rates were highest in the Axx condition (97.5%), which involved one response, followed by the two conditions that involved two responses, ABx (93.6%) and AAx (92.4%), and lowest in the three-response ABA condition (91.7%). Importantly, there was no difference in accuracy between the individual (AAx) and joint (ABx) two-object naming conditions ($B = -.45$, $SE = .30$, $z = -1.48$).

Naming latency. As shown in Figure 2 (Panel A), naming condition had a large effect on latencies to begin naming the first object. Most strikingly, naming latencies were slower in the AAx condition than in all other conditions. As expected, Participant A was faster to initiate a one-object response (Axx) compared with a two-object response (AAx; $B = -.14$, $SE = .01$, $t = -11.46$). Crucially, Speaker A's naming latencies in the ABx condition were also faster than those in the AAx condition ($B = -.14$, $SE = .01$, $t = -11.69$) and did not differ from those in the Axx condition ($B = -.00$, $SE = .01$, $t = -.06$). When Speaker A named both the first and third object (ABA), naming latencies were somewhat slower than when they named only the first object, whether in the Axx condition ($B = .03$, $SE = .01$, $t = 2.51$) or ABx condition ($B = -.03$, $SE = .01$, $t = -3.46$), but faster than when they named the first and second object (AAx; $B = .11$, $SE = .01$, $t = 8.74$). Relatedness did not have a significant effect on first-object naming latencies in any of the naming conditions ($ts < 2$).

The naming latencies on filler trials, where Speaker B named the first object, showed a similar pattern. Naming latencies were longest in the BBx condition (908 ms, $SD = 94$ ms) and differed significantly from the latencies in the Bxx condition (768 ms, $SD = 73$ ms; $B = .16$, $SE = .01$, $t = 12.24$) and the BAx condition (824 ms, $SD = 77$ ms; $B = .10$, $SE = .02$, $t = 4.02$). Interestingly, the difference between the latter two conditions (Bxx and BAx), which failed to reach significance for Speaker A, was significant here ($B = .07$, $SE = .02$, $t = 2.90$). When Speaker B named both the first and third object (BAB; $M = 863$ ms, $SD = 87$ ms), naming latencies were again slower than when they named only the first object, whether in the Bxx condition ($B = .11$, $SE = .02$, $t = 4.44$), or BAx condition ($B = .04$, $SE = .01$, $t = 3.59$), but they were faster than when Speaker B named the first and second object (BBx; $B = -.05$, $SE = .02$, $t = 2.24$). Relatedness did not have a significant effect on first-object naming latencies in any of the naming conditions ($ts < 2$). (To achieve convergence, the models assessing naming latencies on filler trials did not include random slopes for the effect of relatedness and no by-item random slopes for the effect of naming condition).

As described, participants were asked not to leave silent pauses between object names. To assess how well they succeeded, we

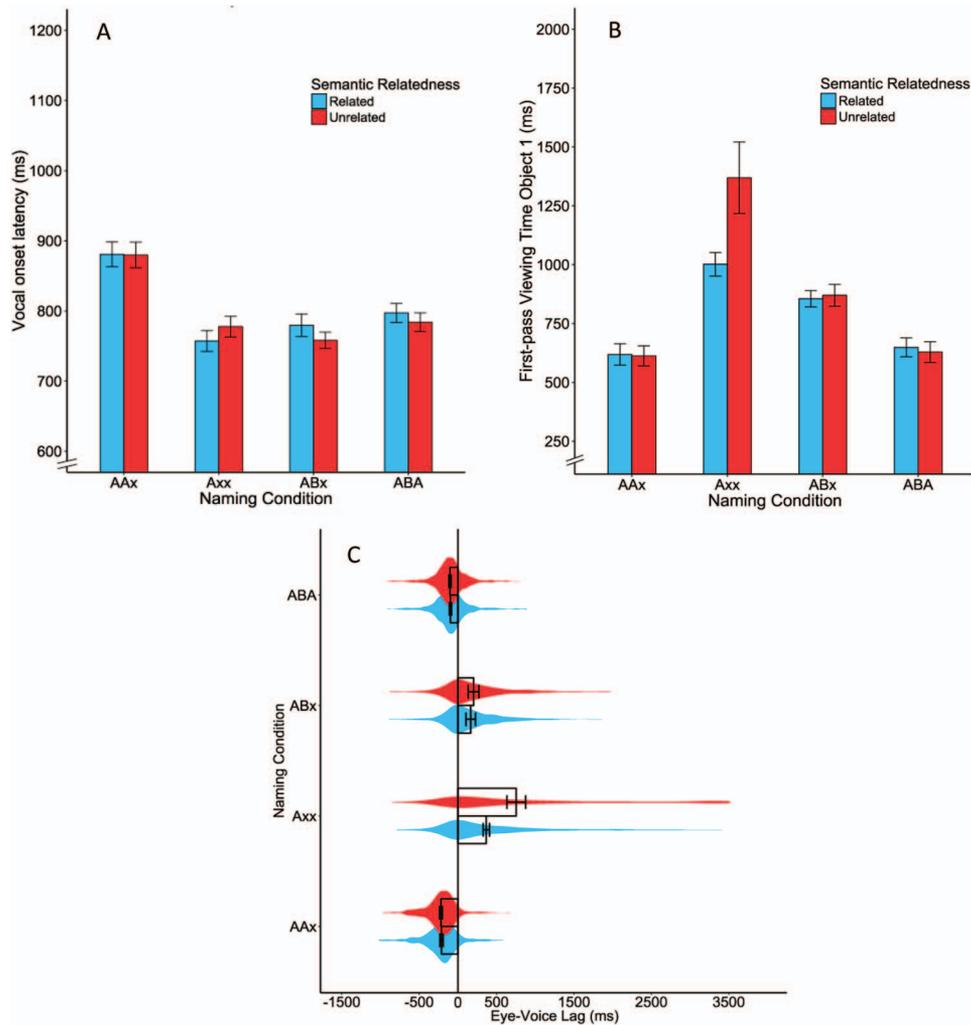


Figure 2. Latencies to begin naming the first object (Panel A) and first-pass viewing times on the first object (Panel B) and eye-voice lead (EVL; Panel C) by naming condition in Experiment 1. EVL is computed as the timing of the gaze shift away from the first object relative to the vocal onset of the first object name. Zero on the x-axis represents the vocal onset of the first object name. If the eye-voice lag is negative, the gaze shift occurred before the vocal onset. If the eye-voice lag is positive, the gaze shift occurred after the vocal onset. All error bars represent normalized within-participant 95% confidence intervals (Loftus & Masson, 1994). For EVL, within-participant error bars were computed separately within the AxAx/ABA conditions (two vocal responses from Speaker A, mostly negative EVL values) and the AxAx/ABx conditions (one vocal response from Speaker A, mostly positive EVL values). See the online article for the color version of this figure.

analyzed the proportion of trials on which a silent pause could be detected between naming the first and second object and the average duration of such pauses if present. The results are presented in Table 2. Utterances from Speakers A and B overlapped on a small proportion of trials (<1% of all trials). As mentioned, trials on which the overlap caused a disruption were not included in the analysis. For the purpose of the current analysis, the remaining trials on which the onset of one speaker's response slightly overlapped with the offset of the other speaker's response were coded as having no silent pause between responses. On most trials there was a short silent pause between the first and second object name, and these interword pauses occurred more often for between- than within-speaker transi-

tions. Comparing the two two-response conditions, the average proportion of trials with a silent pause between the first and second object name was higher in the ABx than in the AxAx condition ($B = 2.05$, $SE = .67$, $z = 3.04$), and, if present, average pause duration was longer for between-speaker pauses than within-speaker pauses ($B = .18$, $SE = .07$, $t = 2.64$). However, in the ABA condition, between-speaker transitions between the first and second word more closely resembled the within-speaker transitions in the AxAx condition. The average proportion of trials with a silent pause between the first and second word in the ABA condition was not significantly different from the AxAx condition ($B = 1.00$, $SE = .59$, $z = 1.70$), and neither was the average pause duration ($B = .10$, $SE = .06$,

Table 2
Proportion of Trials With Silent Pauses Between the First and Second Object Name and Duration of Pauses if Present in Experiment 1

Trial	Proportion trials with silent pause <i>M (SD)</i>	Pause duration if present (ms) <i>M (SD)</i>
Naming condition		
Axx		
AAx	.82 (.25)	266 (117)
ABx	.91 (.17)	294 (107)
ABA	.87 (.23)	276 (112)

$t = 1.57$). Relatedness did not have a significant effect on the proportion of trials with silent interword pauses ($z_s < 2$) or silence duration ($t_s < 2$).

Fixation patterns and viewing time. Fixations patterns are presented in Table 3. As noted, we only included trials on which participants looked at the first object at trial onset. As expected based on results of numerous earlier multiple-object naming experiments, the participants looked at the objects in the order of naming. In the AAx and in the ABx conditions, participants moved their eyes from the first object directly to the second object on more than 98% of trials. Thus, with respect to spatial fixation patterns, these two conditions were equivalent. In the ABA condition, the participants also most often shifted their gaze from the first to the second object, but on 29% of the trials they skipped the second object and moved straight to the third object. Such skipping rarely occurred in the ABx and AAx conditions (less than 2% of the trials). A logistic regression model confirmed that the probability of first-pass skipping the second object was higher in the ABA than in the ABx condition ($B = 4.35$, $SE = .83$, $z = 5.24$), and the AAx condition ($B = 8.28$, $SE = .365$, $z = 2.27$). Relatedness did not affect skipping rates ($z_s < 1.5$). (To achieve convergence, random slopes for the effect of relatedness and the random slope for the effect of naming condition by items were removed.)

First-pass viewing times for the first object are shown in panel B of Figure 2. To achieve convergence, the by-item random slope for the effect of naming condition was removed from the models assessing viewing time. In the Axx condition, only the first object was task relevant; hence it is not surprising that in this condition participants looked at this object longer than in the remaining conditions. Interpreting the duration of the viewing time to the first object in this condition is problematic since participants had no other task-relevant target to move their eyes to. Turning to the remaining conditions, viewing times for the first object were longest in the ABx condition, followed by the ABA condition, $ABx > ABA$ ($B = -.26$, $SE = .02$, $t = -12.43$), and shortest in the AAx condition, $ABA > AAx$ ($B = -.04$, $SE = .02$, $t = 2.18$). The effect of relatedness was significant only in the Axx condition, such that first-object viewing times were longer in the unrelated compared to the related condition ($B = -.22$, $SE = .03$, $t = -8.68$).

We also examined the first-pass viewing times for the second object. These gaze durations are difficult to interpret in all conditions except for the ABA condition since no other objects were

named after the second object, and participants had no incentive to move their eye gaze to a new location. Still, it is worth noting that Speaker A's average viewing time of the second object in the ABA condition was only 380 ms ($SD = 151$ ms), which is approximately half of the viewing time seen for the first object of the present study and observed for objects of this complexity in earlier studies (e.g., Levelt & Meyer, 2000; Meyer, 1996; Meyer et al., 1998).

Finally, we analyzed the *eye-voice lag* (EVL), defined as the time between the gaze shift away from object 1 and the onset of speech (see Panel C of Figure 2). EVL is an indicator of the coordination of the participants' speech onset and eye gaze. When the voice onset latency is longer than the first pass viewing time for object 1, EVL is negative. When the voice onset latency is shorter than first-pass viewing time for object, EVL is positive.

As EVL could include zero or negative values, the statistical analyses were based on raw rather than log-transformed data. EVLs with absolute values over 2000 ms were excluded from the analyses. This exclusion affected 2.9% of trials, all of which were cases in the Axx condition when Speaker A shifted their gaze away from the first object more 2,000 ms after the onset of the object name. As shown in Figure 2, when Speaker A named two objects in the trial (AAx and ABA conditions), EVL was most often negative, meaning Speaker A shifted their gaze away from Object 1 before the onset of its name. When Speaker A named one object in the trial (Axx and ABx conditions), EVLs were most often positive, meaning Speaker A shifted their gaze away from object 1 after the onset of its name. A model comparing EVLs in the AAx and ABA conditions (sum-to-zero coded: $-.5$) and Axx and ABx conditions (sum-to-zero coded: $.5$) showed that the effect of the number of Speaker A's responses on EVL was significant ($B = 401.02$, $SE = 21.06$, $t = 19.04$). Moreover, comparisons between individual conditions showed that EVLs in the AAx condition (mostly negative) were significantly different from those in the ABx condition (mostly positive); $B = -403.41$, $SE = 21.75$, $t = -18.55$. (To achieve convergence, random slopes for the effect of relatedness were not included in this model). Comparison of the two one-response conditions showed that EVLs in the ABx condition (coded: $-.5$) were smaller (gaze shift occurring closer to the vocal onset) than in the Axx condition (coded: $.5$; $B = 123.13$, $SE = 27.72$, $t = 4.44$). Comparison of the two two-response conditions showed that EVLs in the AAx condition (coded $-.5$)

Table 3
Fixation Patterns by Naming Condition in Experiment 1

Trial	Object 1-2 <i>M (SD)</i>	Object 1-3 <i>M (SD)</i>	Object 1 only <i>M (SD)</i>
Naming condition			
Axx	90.1 (15.2)	2.4 (3.9)	7.5 (14.3)
AAx	99.5 (1.8)	.5 (1.8)	0
ABx	98.2 (5.3)	1.0 (2.1)	.8 (4.1)
ABA	71.4 (24.6)	28.6 (.25)	0

Note. Percentage of trials on which Speaker A moved their eyes directly from the first object to the second object (Object 1-2), from the first object directly to the third object (Object 1-3), or stayed on the first object for the duration of the trial (Object 1 only). This classification of fixation patterns does not take into account any eye movements that occurred after the first between-object saccade.

were longer (gaze shift occurring earlier relative to the vocal onset) than in the ABA condition (coded .5; $B = 117.38$, $SE = 11.6$, $t = 10.12$).

Discussion

Experiment 1 investigated how speakers corepresent each other's speech planning operations during a joint naming task with consecutive responses. High accuracy rates and short interturn pauses indicated that participants could successfully perform the joint task of producing fluent, multiword utterances. Average pause durations between the first and second object name were under 300 ms in all naming conditions and thus well within the range of interturn pause durations observed by [Stivers and colleagues \(2009\)](#). Moreover, silent pauses occurred only slightly more often on two-speaker ABx trials than on one-speaker AAx trials (89% vs. 82% of the trials) and were only slightly longer in the former than in the latter condition (285 ms vs. 266 ms). Thus, the participants' utterances were generally well coordinated.

The results in the conditions where Speaker A was the only participant who spoke (Axx and AAx) replicate earlier findings on multiple object naming in single-speaker situations: naming latencies were longer when speakers were planning to name two objects relative to one object (e.g., [Griffin, 2001](#); [Levelt & Meyer, 2000](#); [Meyer, 1996](#); [Meyer et al., 1998](#)). The main objective of the current experiment was to compare measures of speech planning in the individual naming conditions (Axx and AAx) to the joint naming conditions (ABx and ABA). We discuss the results in the two joint-naming conditions separately, starting with the ABx condition.

According to the corepresentation hypothesis, interlocutors develop a shared representation of the joint task. Consequently, Speaker A should include the second object in their speech planning in the same way in the ABx and the AAx condition, resulting in similar speech onset latencies in these two conditions. Moreover, the average latencies for both conditions should be longer than the average in the Axx condition. These predictions were not borne out: Speaker A's naming latencies were shorter in the ABx than in the AAx condition, and the latencies in the ABx and Axx condition did not differ significantly from each other. This pattern indicates that Speaker A did not include B's naming task in their own speech plan. Speaker B's latencies showed a slightly different pattern: Latencies were shorter in the BBA condition than in the BBx condition; however, for this speaker, significantly longer latencies were observed in the BBA than in the BBx condition. This pattern indicates that Speaker B paid some attention to Speaker A's object when planning their own utterance, but did not fully incorporate this object into their own speech plan.

The fixation patterns are best understood in combination with the naming latencies. When Speaker A was to name two objects (AAx condition), viewing times on the first object were shorter than in any other condition and EVL was typically negative, meaning the gaze shift occurred before the vocal onset. When Speaker A was to name one object (Axx condition), viewing times on the first object were very long compared to the other naming conditions and EVL was typically positive, meaning the gaze shift occurred after the vocal onset. These results suggest Speaker A in the AAx condition began planning the second object name before initiating the first object name. This is consistent with earlier

findings on the coordination of eye gaze and speech in single-speaker naming tasks (e.g., [Gordon & Hoedemaker, 2016](#); [Griffin, 2001](#); [Levelt & Meyer, 2000](#); [Meyer, 1996](#); [Meyer et al., 1998](#); [Roelofs, 2007, 2008](#)).

Crucially, in the joint-naming condition (ABx), Speaker A's EVL was also largely positive and quite variable, more closely resembling the pattern in the one-object (Axx) than the two-object (AAx) single-speaker condition. In other words, Speaker A's pattern of eye-speech coordination reflected the number of objects they were planning to name themselves (their individual task) rather than the number of objects Speaker A and B were naming together (the shared task). First-object viewing times in the ABx condition were shorter than in the Axx condition, but longer than in the AAx condition. This suggests that Speaker A had more interest in inspecting the second object when it was to be named by Speaker B than when it was not named by either speaker, but was never as eager to inspect the object as when they were planning to name it themselves. Most importantly, they typically did not inspect the second object before beginning to speak, as they did when they had to name that object.

Behavior in the three-object joint-naming condition (ABA) differed from both the ABx and AAx conditions. The ABA condition was included to investigate whether corepresentation was encouraged when it might be self-serving for the individual speaker. Unlike the ABx condition, the ABA condition required Speaker A to respond to B. Therefore, they had to time their utterance to the ending of B's turn. Representing Speaker B's operations could help Speaker A estimate the timing of B's utterance, allowing them to time their own subsequent response with greater precision. Naming latencies were longer in the ABA condition than in the ABx condition, but not as long as in the AAx condition. In addition, in the ABA condition, the gaze shift away from the first object typically occurred before the onset of the first name (negative EVL), although not as early (relative to speech onset) as in the AAx condition. These findings might be interpreted to mean that Participant A engaged in some planning of the second object name even though they were not assigned to name it, supporting the corepresentation hypothesis. In addition, this pattern of naming latencies and eye movements in the ABA condition falling in between that of the AAx and the ABx conditions is consistent with a partial corepresentation account ([Gambi et al., 2015a](#)). According to this account, interlocutors corepresent their partner's task using some, but not all, of the same mechanisms engaged in their own object-naming process.

However, an alternative interpretation of the observed behavior in the ABA condition is that the slower naming latencies and negative EVLs reflect the fact that in this condition Speaker A also named the third object. This interpretation is supported by the observation that Speaker A moved their eyes directly from the first to the third object, thereby first-pass skipping the second object, on 29% of ABA trials. In the other naming conditions, the second object was hardly ever skipped, and previous research has also found that speakers rarely skip objects they intend to name ([Van der Meulen, Meyer, & Levelt, 2001](#)). Therefore, it is not clear how to interpret this finding. If Speaker A represented Speaker B's task in the ABA condition as if it were their own task, they should not have skipped the second object as frequently as they did. In addition, average first-pass viewing time for the second object in the ABA condition (380 ms) was considerably shorter than the

viewing time of around 600 ms that would be expected if Speaker A was planning to name the object (Meyer et al., 1998). However, if Speaker A regarded the second object in the ABA condition as being completely irrelevant to their task, we might expect them to have skipped this item even more consistently. Interpretation of these results is further complicated by the fact that the second object was located in between the first and third object. That is, some of the first-pass fixations on the second object may be the result of the viewer intending to skip the object but undershooting this long saccade and unintentionally fixating the object for a short time (Kapoula & Robinson, 1986; McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; Rayner, 1998). The main goal of Experiment 2 was to clarify how participants processed the second objects in ABA sequences.

Semantic relatedness between the first and second object name did not show a clear pattern of effects across naming conditions. The two objects were presented close enough together to allow for some extrafoveal processing of the second object during the inspection of the first object (Malpass & Meyer, 2010; Morgan & Meyer, 2005). Therefore, one might expect to see interference, that is, longer naming latencies and viewing times for the first object in the related than in the unrelated condition, in particular in the AAX and perhaps in the ABx conditions, where the second object was task-relevant (Aristei et al., 2012; Freedman et al., 2004; Schotter, Ferreira, & Rayner, 2013). However, the naming latencies did not show a relatedness effect in any of the three naming conditions. This suggests that participants initially strongly prioritized the processing of the first object, and any extrafoveal processing of the second object was too slow or inconsistent to affect the naming latencies for the first object (Jescheniak et al., 2014; Matushanskaya et al., 2016). An effect of relatedness on the first-pass viewing time for the first object was only seen in one naming condition, the Axx condition, with participants' eye gaze moving earlier to related than to unrelated second objects. As Figure 2 shows, the gaze shift typically occurred well after the onset of speech. Thus, the related objects attracted the participant's visual attention earlier than the unrelated ones, but only after the first object name had been prepared. Why the processing of the extrafoveal object was less efficient in the present experiment than in

earlier studies is unclear. One possibility is that the objects used here were harder to identify or to name than those in the earlier studies so that during the processing of the first object less processing capacity was available for the processing of the second object (Malpass & Meyer, 2010). With respect to our main question—how the participants processed the objects to be named by their partners—the semantic relatedness manipulation turned out to be uninformative. Therefore, only unrelated objects were used in Experiment 2.

Experiment 2

Experiment 2 further investigated speech planning in the ABA condition. In Experiment 1, Speaker A's behavior in this condition suggested they might have engaged in some degree of corepresentation in order to perform the shared task. However, the spatial arrangement of the objects made the fixation patterns difficult to interpret. In addition, the ABA condition was the only condition in which either speaker named the third object, thus lacking a proper comparison for fixation durations on the second object. To address these issues, Experiment 2 tested the ABA condition again, but the spatial arrangement of the objects was changed to a triangular configuration (see Figure 3) so that the second object was no longer in between the first and third. With this arrangement, Speaker A could move their eyes directly from the first to the third object without needing to skip over the second. As a result, observed fixations on the second object are more likely to be intentional rather than a consequence of saccadic error. In addition, we included two new control conditions. In the AAA condition, Speaker A named all three objects. In the AxA condition, they named the first and third object, skipping the second object, which was not named by either speaker. Thus, they simply named two objects in a row just as in the AAX condition in Experiment 1.

The main question was whether fixation patterns in the ABA condition would resemble those in the AxA or the AAA condition. In the AxA condition, the second object was neither task-relevant nor located en route to the third object and should therefore receive very few first-pass fixations. In the AAA condition, the second object should almost always be inspected immediately after the

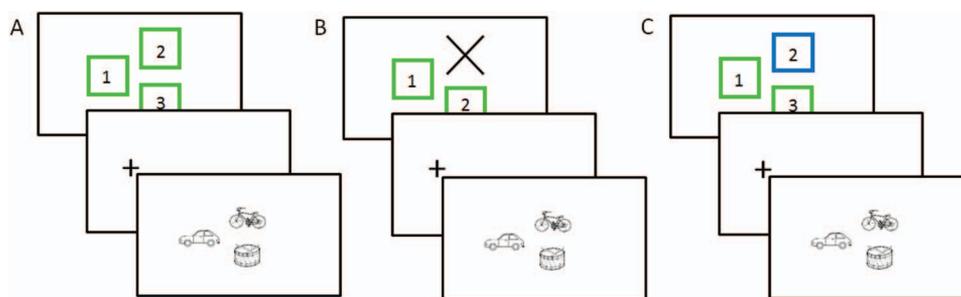


Figure 3. Example of the stimulus presentation in Experiment 2. The boxes preceding the pictures instructed Speaker A (green (shown as grey) boxes) and Speaker B (blue (shown as black) box) which objects they should name. Panels A, B, and C show the instruction screens for the conditions AAA, AxA, and ABA, respectively, in the clockwise naming direction. The filler conditions in which the A–B assignment was reversed are not depicted. The object numbers are included here to illustrate naming order, but were not present in the actual stimulus displays. For the counterclockwise naming direction, assignment of the objects to be named second and third was reversed. See the online article for the color version of this figure.

first object (i.e., receive a first-pass fixation rather than be skipped). This would replicate the pattern seen in many earlier multiple object naming experiments. The ABA condition was the only joint-naming condition in this experiment. If Speaker A corepresents Speaker B's task on some of the trials, they should, as in Experiment 1, occasionally direct their gaze from the first to the second object, as if they were going to name this object themselves. In addition, if Speaker A retrieves the name of the second object even when it is Speaker B's task to name it, their first-pass viewing times on this object should be similar in the AAA and ABA conditions. By contrast, if the first-pass fixations on the second object in Experiment 1 resulted mainly from saccade errors rather than corepresentation of Speaker B's task, participants should now be as unlikely to look at the second object in the ABA as in the AxA condition.

In the present experiment, the speakers had to name either two or three objects on all trials. Since earlier studies have shown that speakers in multiple object naming tasks rarely prepare more than two object names before speech onset (e.g., Gordon & Hoedemaker, 2016; Griffin & Bock, 2000; Meyer, Wheeldon, Van der Meulen, & Konopka, 2012), we did not expect large differences between the latencies to begin speaking in the three conditions. Similarly, since in all conditions at least one other object was to be named after the first one, the viewing times for the first object were not expected to vary much as a function of naming condition.

Method

Participants. A total of 24 pairs of native speakers of Dutch participated in the experiment, none of which had participated in the previous experiment. Sixteen pairs consisted of two female participants, six pairs consisted of one male and one female participant and the remaining two pairs consisted of two male participants. Average participant age was 23.8 years (range = 19–49 years). Ethical approval to conduct the study was given by the Ethics Board of the Social Sciences Faculty of Radboud University.

Materials and design. The stimulus materials were the same as in Experiment 1 with the exception that the semantic relatedness manipulation was dropped from the design, so that each list contained only trials with unrelated objects in all three positions. These were the same unrelated object triplets as used in Experiment 1, except for six objects that had shown relatively low naming accuracy in Experiment 1 and were replaced (see the Appendix and Table 1). As in Experiment 1, in the single object naming experiment (see the Materials section for Experiment 1), the naming latencies for the objects in first position were shorter than for the objects in the second position ($B = 65.52$, $SE = 19.12$, $t = 3.43$). The items were presented in a 3 (naming condition: ABA, AAA, and AxA) \times 2 (naming direction: clockwise vs. counterclockwise; see the following text) mixed design. Naming condition was manipulated within participants. The ABA condition was identical to the ABA condition used in Experiment 1, such that Speaker A named the first object, followed by Speaker B naming the second object, followed by Speaker A naming the third object. In the AxA condition Speaker A named the first and third object (without pausing in between) and nobody named the second object. In the AAA condition, Speaker A named all three objects in 1–2–3 order. Filler items had the speaker roles reversed,

resulting in the BAB, BBB, and BxB conditions. Critical and filler items were distributed across three counterbalanced lists, such that each position-1 object appeared twice per list, once in a critical trial (AAA, ABA, or AxA) and once in a filler trial (BBB, BAB, or BxB), resulting in 160 trials per list. Across lists, each object appeared once in each naming condition and trial type (critical and filler) combination and each list contained an approximately equal number of items (i.e., 26 or 27 items) in each trial type and each naming condition. Six practice trials using objects that did not appear elsewhere in the experiment were added to the start of each list. Data from the practice items were not analyzed.

In the clockwise naming direction, the second object was presented in the top right position of the triangular spatial arrangement and the third object in the bottom right position. In the counterclockwise direction this was reversed. Naming direction was manipulated between participant pairs. The intention had been to test 12 pairs in each direction condition, but due to an experimenter error, 10 pairs were tested in the clockwise and 14 in the counterclockwise condition. Since the analyses showed that direction did not affect the results in any way, no replacement participants were tested.

Apparatus and procedure. The equipment and experimental setup were the same as in Experiment 1. Instructions to each participant pair included only the appropriate naming direction (clockwise or counterclockwise).

Analysis. The analysis strategy was largely the same as in Experiment 1. 10.3% of the trials were excluded because of naming errors. These were trials where participants named one or more objects incorrectly (57% of the error trials) and/or hesitated (47% of the error trials) and/or failed to follow the instructions by omitting an object (14% of the error trials), naming objects in the incorrect order (7% of the error trials), or initiating a turn too early so that the naming sequence failed (2% of the error trials). Trials on which Speaker A did not start by fixating the first object were removed from the analysis (1.9% of correct critical trials). Trials where first object naming latencies or first-pass viewing times differed more than 2.5 standard deviations from the mean for that particular naming condition were trimmed from the analysis (4.4% of remaining correct trials). Naming latencies and eye movement measures were analyzed using linear and logistic mixed-effects models applied in the same way as in Experiment 1. Unless otherwise mentioned, all models included random intercepts for participants and items and random slopes for the by-participant and by-item effects of naming condition. All models initially included naming direction as a fixed effect predictor (sum-to-zero coded). However, as naming direction never had a significant effect on the measures of interest, this predictor was dropped from all models and will not be discussed further.

Results

Naming accuracy. Average trial accuracy across pairs of participants was 89.7% (range = 75.0%–97.5% correct trials). Accuracy was higher when only Speaker A spoke during a trial compared with when the speakers were required to alternate: accuracy rates in the ABA condition (86.5%) were lower than in the AxA (91.7%; $B = .79$, $SE = .32$, $z = 2.42$) and AAA (90.8%) conditions ($B = .73$, $SE = .30$, $z = 2.44$). Accuracy rates in the

AAA and AxA conditions did not differ significantly ($B = .05$, $SE = .36$, $z = .15$).

Naming latency. Speaker A's naming latencies for the first object are presented in Panel A of Figure 4. Somewhat unexpectedly, naming latencies were longer when Speaker A was preparing to name three objects (AAA) compared with two objects in a row (AxA; $B = .05$, $SE = .01$, $t = 4.37$). Crucially, Speaker A's naming latencies in the joint-naming condition (ABA) were shorter than in either of the two conditions in which A was the only speaker: $ABA < AAA$ ($B = -.17$, $SE = .02$, $t = -9.74$; $ABA < AxA$ ($B = .12$, $SE = .02$, $t = 7.51$).

The naming latencies observed on filler trials, when Speaker B named the first object, followed the same pattern, being longest in

the BBB condition ($M = 970$ ms, $SD = 92$), followed by the BxB condition ($M = 942$ ms, $SD = 94$) and the BAB condition ($M = 804$, $SD = 67$). Again, the average latency in the joint naming condition (BAB) was significantly shorter than the averages in the BBB condition ($B = .19$, $SE = .02$, $t = 10.5$) and in the BxB condition ($B = 0.16$, $SE = .02$, $t = 10.6$). (To achieve convergence, this model did not include a random slope for the effect of naming condition by items.)

As in Experiment 1, we analyzed the probability and durations of interword silent pauses as measures of speaker coordination (see Table 4). As mentioned, trials on which overlapping speech disrupted subsequent responses were removed from the data. Responses by Speaker A and Speaker B overlapped slightly, without

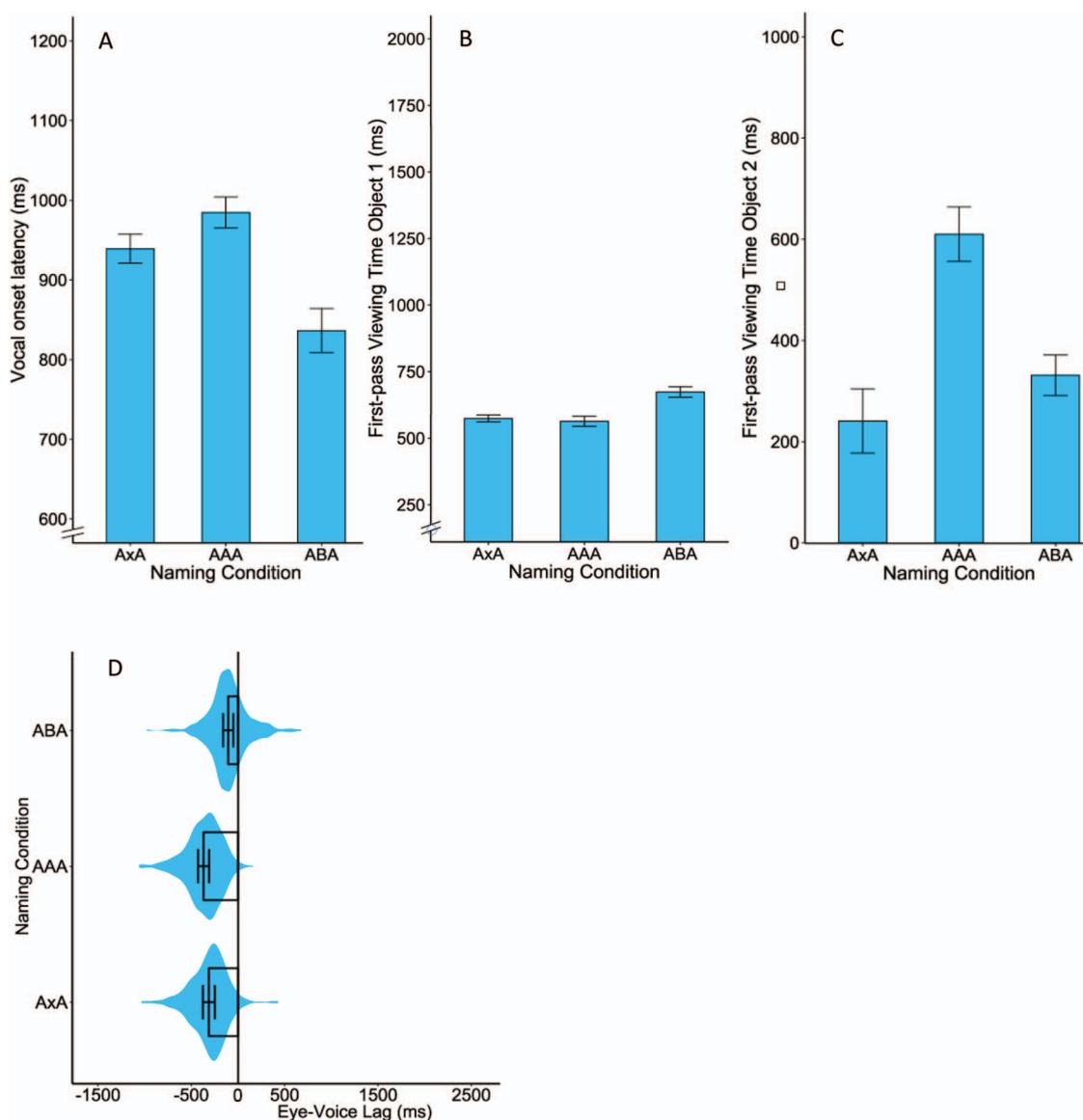


Figure 4. Latencies to begin naming the first object (Panel A), first-pass viewing times on the first (Panel B) and second object (Panel C) and eye-voice lag (EVL; Panel D) in Experiment 2. EVL was computed the same way as in Experiment 1. Error bars represent normalized, within-participant 95% confidence intervals (Loftus & Masson, 1994). See the online article for the color version of this figure.

Table 4

Proportion of Trials With Silent Pauses Between the First and Second Object Name and Between the Second and Third Object Name and Duration of Pauses if Present (Means and Standard Deviations) in Experiment 2

Trial	Proportion trials with silent pause between first and second object <i>M (SD)</i>	Pause duration if present (ms) <i>M (SD)</i>	Proportion trials with silent pause between second and third object <i>M (SD)</i>	Pause duration if present (ms) <i>M (SD)</i>
Naming condition				
AAA	.78 (.18)	232 (71)	.79 (.19)	211 (70)
AxA	.82 (.20)	296 (126)		
ABA	.93 (.13)	231 (66)	.88 (.19)	241 (84)

disrupting the sequence, on a small proportion of trials (<1%). These trials were coded as including no pause. A short silent pause between the names of the first- and second-named object was observed on most trials, but more often for between- than within-speaker transitions. In the ABA condition, a silent pause between Speaker A's naming of the first object and Speaker B's naming of the second object was observed significantly more often than in the AxA condition ($B = -1.68$, $SE = .61$, $z = -2.78$) and the AAA condition ($B = -2.04$, $SE = .61$, $z = -3$). Proportion of trials with a silent pause did not differ for the two within-speaker conditions (AxA vs. AAA; $B = .36$, $SE = .46$, $z = .77$). However, when silent pauses were present, the average durations of the pauses did not differ as a function of speaker transition: The average pause duration in the ABA condition was highly similar to the average pause duration in the AAA condition ($B = -.11$, $SE = .06$, $t = -1.76$) and in the AxA condition ($B = .10$, $SE = .09$, $t = .13$).

Results were similar for the transition between the second and third object name. The probability of observing a silent pause between the second and third word differed for within- and between-speaker transitions. Silent pauses in this position were observed more often between speakers than within speakers ($B = 1.49$, $SE = .53$, $z = 2.83$). Again, when a silent pause between the second and third word was present, there was no difference in average duration of the silence in the AAA and ABA condition ($B = -.09$, $SE = .11$, $t = -.85$). All in all, the results indicate that speakers successfully coordinated their consecutive responses.

Fixation patterns and viewing time. The mean proportions of trials with particular viewing patterns in each naming condition across participants is presented in Table 5. As noted, analyses

included only trials on which participants fixated the first object at the onset of the trial. In the AAA condition, Speaker A virtually always fixated all three objects in order and, as expected, the second object was never skipped. In the AxA condition, Speaker A almost always moved their eyes directly from the first to the third object, resulting in a first-pass skip of the second object on 98% of trials. This was expected as well since the second object was not named by either speaker and thus was not relevant to the task. The ABA condition showed an intermediate pattern, with Speaker A fixating the second object on the first pass on approximately 25% of trials and skipping the object on the remaining 75% of trials. Statistical analysis confirmed that first-pass skipping behavior in the ABA condition differed significantly from both the AxA and the AAA condition. An intercept-only model showed that the average skipping rate in the ABA condition was significantly higher than zero (representing the AAA condition; $B = 1.56$, $SE = .25$, $z = 4.43$). Comparison of the ABA and AxA condition showed that skipping rates in the ABA condition were significantly lower than in the AxA condition ($B = -2.12$, $SE = .51$, $z = 4.17$).

First-pass viewing times on object 1 (see Panel B of Figure 4) were significantly longer in the ABA condition than in the AAA condition ($B = -.16$, $SE = .02$, $t = -7.59$) and the AxA condition ($B = -.14$, $SE = .02$, $t = -8.58$). Viewing times in the AAA and AxA condition did not differ ($B = .02$, $SE = .02$, $t = .89$). Analysis of first-pass viewing times on the second object is complicated by the fact that in some naming conditions (ABA and especially AxA), it received very few first-pass fixations, resulting in an uneven number of observations across cells. Nonetheless, the differences in first-pass viewing time as a function of naming condition were quite clear (see Panel C of Figure 4). Average first-pass viewing time on the second object in the AAA condition ($M = 610$ ms, $SD = 77$ ms) was consistent with previous observations during single-speaker object naming tasks (e.g., Indefrey & Levelt, 2004). In contrast, second-object viewing times in the ABA condition were much shorter ($M = 332$ ms, $SD = 118$ ms), $ABA < AAA$ ($B = -.57$, $SE = .08$, $t = -7.15$). Interestingly, viewing times were slightly longer in the ABA condition than in the AxA condition ($M = 241$ ms, $SD = 75$ ms) and the $AxA < ABA$ ($B = -.29$, $SE = .14$, $t = 2.02$), meaning Speaker A spent more time inspecting the second object when it was to be named by Speaker B than when it was not named at all.

Consistent with the fact that Speaker A named more than one object in all three conditions, EVL was typically negative regardless of naming condition (see Panel D of Figure 4). In the ABA

Table 5

Fixation Patterns by Naming Condition in Experiment 2

Trial	Object 1–2 <i>M (SD)</i>	Object 1–3 <i>M (SD)</i>
Naming condition		
AAA	100 (.0)	0
AxA	2.4 (3.0)	97.6 (3.0)
ABA	24.9 (25.3)	75.1 (25.3)

Note. Percentage of trials on which Speaker A moved their eyes from the first to the second object (Object 1–2) and from the first directly to the third object (Object 1–3). This classification of fixation patterns does not take into account any eye movements that occurred after the first between-object saccade.

condition, EVL was shorter (closer to the vocal onset) than in the AAA condition ($B = -268.81$, $SE = 17.46$, $t = -15.40$) and shorter than in the AxA condition ($B = -211.56$, $SE = 16.98$, $t = -12.46$). (To achieve convergence, this model did not include random slopes for the effect of naming condition by items.)

Discussion

The goal of Experiment 2 was to further investigate whether speakers corepresented each other's speech planning processes in an interactive joint naming task. Specifically, we asked whether and for how long Speaker A would inspect objects that were part of the shared naming task, but that were to be named by Speaker B rather than Speaker A themselves. As in Experiment 1, the average duration of silent pauses between words was under 300 ms for both intra- and interturn pauses, matching the typical duration of interturn pauses in normal dialogue (Stivers et al., 2009). Silent pauses between words occurred more often in the ABA condition (pause between first and second word: 93% of trials, pause between second and third word: 88% of trials) than in the AAA condition (pause between first and second words: 78% of trials, pause between second and third word: 79% of trials), but overall the timing of the utterances in the two conditions was very similar. These results suggest that the speakers successfully coordinated their responses in time. Nonetheless, the remaining results suggest that this coordination did not rely on Speaker A corepresenting Speaker B's task as if it were their own. Instead, we observed evidence that this level of coordination was achieved while Speaker A prioritized their own naming task.

According to the corepresentation account, speakers should incorporate their own and their task partner's items into their speech planning process. Consequently, Speaker A should attend to each object equally regardless of who was meant to overtly name it, resulting in similar fixation rates and viewing times for the second object across the ABA and AAA conditions. Consistent with the fact that Speaker A named more than one object in all three conditions, EVL was typically negative regardless of naming condition (see panel D of Figure 4). As expected, Speaker A always fixated the second object before the third one when they were naming all three objects (AAA), and almost always skipped the second object when nobody was naming it (AxA). Crucially, in the ABA condition, Speaker A (first-pass) fixated the second object only about 25% of the time. Thus, they did not attend to the second object as consistently as when they named it themselves. These results also suggest that a substantial proportion of the first-pass fixations on the second object in the ABA condition in Experiment 1, in which the fixation rate for this object was 75%, were the result of fixation error related to the spatial arrangement of the objects.

The analysis of the speech onset latencies showed that Speaker A was faster to initiate their vocal response in the joint-naming condition (ABA) than in either of the conditions in which they were the only speaker (AAA and AxA). The same pattern was seen for Speaker B, who spoke first on filler trials. These findings suggest that rather than incorporating the partner's task of naming the second object into their own speech planning, the first speaker took advantage of the fact that they were not responsible for naming this object. Knowing that Speaker B would name the second object provided Speaker A time to prepare the third object

name once naming of the first object had been initiated. As a result, Speaker A could start naming the first object earlier in the joint than in the individual naming conditions.

Nonetheless, speakers did not appear to be completely indifferent to their partner's task. Speakers were more likely to fixate the second object when it was about to be named by their partner than when nobody was to name that object. This observation is consistent with the established finding that listeners tend to fixate objects they expect to be mentioned (Altmann & Kamide, 1999) and with the observation that when two speakers discuss a shared visual representation, their eye movements tend to become coupled (Richardson, Dale, & Kirkham, 2007). However, the current result does not provide evidence for a strong interpretation of the corepresentation account, which would predict speakers to fixate their partner's objects with as much regularity as their own objects. In our experiment, Speaker A always first-pass fixated their own objects, but looked at Speaker B's objects on the first-pass much less often (approximately 25% of trials). Moreover, Speaker A's average first-pass viewing time for Speaker B's objects was shorter than the viewing time for their own objects, averaging 332 ms compared with 610 ms. In a time period of 332 ms, Speaker A may be able to recognize the object and perhaps retrieve the appropriate lemma, but can hardly complete the full process of lexical selection and phonological encoding as if they were going to name the object themselves (Indefrey & Levelt, 2004). As will be further discussed below, these results do not support the hypothesis that speakers represent each other's task in exactly the same way as their own task, but instead point toward partial corepresentation that is, representing some of the components of the partner's task some of the time (see also Gambi et al., 2015a, 2015b). In the following text, we discuss further what partial corepresentation may entail and what its functionality may be.

General Discussion

The model of dialogue as joint action (Gambi & Pickering, 2011; Garrod & Pickering, 2009) proposes corepresentation as a functional mechanism that allows for the fast-paced, seemingly effortless coordination that characterizes most dialogue. According to the model, the shared task goal imposed by dialogue leads interlocutors to represent each other's task share as if it were part of their own task. The goal of the present work was to investigate what it means to corepresent the partner's task share in a joint object-naming task. Pairs of participants took turns naming objects presented in sets of three. This required them to coordinate their responses in time to avoid overlapping speech and long silent pauses between utterances. The task lends itself particularly well to corepresentation since both speakers see the same display and can therefore reliably predict their partner's utterances. We tested the strong prediction that the participants' gaze patterns and speech onset latencies would be the same regardless of whether they prepared to name a given object themselves or anticipated that the partner would do so.

Interturn pause times and high overall accuracy (including low rates of overlap) indicated that speakers efficiently coordinated their utterances during this task. Overall, the average duration of interturn silent pauses was within the range observed for everyday dialogue (Stivers et al., 2009). Moreover, between-speaker pauses were only slightly more frequent than within-speaker pauses and

only in Experiment 1 slightly longer. However, the results of the current study do not indicate that this coordination between speakers was based on full corepresentation. First, latencies to speak and measures of eye-voice coordination do not indicate that speakers incorporated their partner's naming task fully into their own speech plan. Speech planning times and eye-voice lags varied strongly as a function of the number of objects the individual speaker was planning to name themselves, and were much less affected by whether or not the partner also named an object. Second, eye movements clearly indicated that speakers preferably attended to objects they were planning to name themselves compared to objects their partner was planning to name.

In the two-response (ABx) condition, these results are perhaps unsurprising. Once Speaker A had named the first object, their portion of the naming task was over and they technically did not need to be concerned with Speaker B's task. However, this approach to the task is inconsistent with the prediction that Speaker A should behave in accordance with the shared goal of naming both the first and the second object in a fluent manner. We hypothesized that an effect of the shared goal might be more readily observed in the ABA condition because it required Speaker A to respond following Speaker B's naming of the second object, and corepresenting Speaker B's speech planning operations might aid the precise timing of Speaker A's response. In this condition, awareness of the interlocutor's task did influence speaker behavior, in the sense that they were more likely to fixate objects about to be named by their task partner than objects they knew would not be named. However, these fixations were too infrequent and viewing times were too short to indicate that speakers were retrieving the names of their partner's objects as if they were planning to name these objects themselves. Taken together, these results suggest that effective coordination between interlocutors was achieved without full corepresentation. Instead, it appears that speakers prioritized their own task, only opting for partial corepresentation on some of the trials.

The results of the current study differ from earlier work that has found evidence in support of more complete corepresentation. Specifically, in an EEG study Baus et al. (2014) found effects of object-name frequency both when the participants prepared to name an object and when they expected their partner to begin to name the object, indicating that the participants represented the partner's task more fully than the participants in the present study. However, the paradigm used in the study by Baus and colleagues was noninteractive in the sense that participants took turns responding to external stimuli (the pictures on the screen) rather than to each other, and on critical trials they listened to their partner knowing that they would not have to speak at least until the start of the next trial. The absence of a task of their own during their partner's naming trials may have encouraged them to "name along" with the other speaker in internal speech (see also Hoedemaker et al., 2017). In the current study, Speaker A always had to name at least one object on critical trials. As a result, potential corepresentation of Speaker B's task was forced to occur within the same time frame as Speaker A's own internal speech planning process. The results suggest that under these circumstances, speakers prioritize their own speech task over full corepresentation of their partner's operations.

The conclusion that the speakers in the present study engaged in partial corepresentation fits in well with conclusions drawn by

Gambi and colleagues (2015a, 2015b). As discussed in the Introduction, their studies showed that representing a partner's task could interfere with the participant's own speech planning when these processes occurred at the same time. In their object naming study, Gambi et al. (2015b) observed slower naming latencies when speakers were aware that their task partner was also preparing to speak compared with when they knew they were the only one to speak on that trial. Importantly, whether or not the partner named the same or a different picture did not affect the participants' latencies. Thus, they only represented part of the processes involved in naming. In the second relevant study Gambi et al. (2015a) asked participants to prepare to name pictures, which on some trials were replaced by new pictures. These new pictures were to be ignored or named by the participant or by their partner. Participants found it harder to stop naming the initial picture when the partner had to name the new picture than when nobody named it, but not as hard as when they had to name it themselves. Although these results indicated that task performance was affected by an awareness of a partner's task, it was not clear to what extent this applied specifically to simultaneous production or whether the effect would generalize to consecutive turn taking that occurs during dialogue. In the present study we found that speakers were more likely to attend to the objects named by their partner than to task-irrelevant objects. Consistent with the findings obtained by Gambi and colleagues, this result demonstrates that the speakers were aware of the partner's task and that this awareness affected their behavior, but that they did not represent each other's object naming assignment as if it were part of their own task.

Given that the participants appeared to represent their partner's task to some extent and some of the time, one may ask, first, whether and how such partial corepresentation facilitated carrying out the joint naming task and, second, how the task was accomplished in the absence of corepresentation. The answers to these questions depend to some extent on the experimental condition. In the ABx condition of Experiment 1 Speaker A completed their share of the task before Speaker B, and in the BAx condition, Speaker B completed their task before Speaker A. There was no immediate benefit for either speaker's speech planning in representing the partner's task share. In fact, we observed a small but significant interference effect in the BAx compared to the Bxx condition. However, as the participants were engaged in a joint naming task, they may have cared about whether or not the other person named the object correctly. To establish whether or not this was the case, they had to look at the partner's object, and indeed this is what Speaker A did on most trials. These gazes to the partner's object typically occurred after Speaker A had completed their own task. In short, in this condition the primary function of the gazes to the partner's object and the resulting corepresentation of their task probably was to check that all objects were named correctly.

In the ABA condition, Speaker A had to time their second response to coincide as closely as possible with the end of Speaker B's utterance. To do this, some representation of Speaker B's task was necessary. Minimally, Speaker A had to remember in the ABA condition that the partner had to take a turn and name an object, whereas in the AxA condition they could name the two objects in immediate succession. Representing Speaker B's task more specifically than simply remembering that Speaker B would

take a turn may have contributed to accurate timing of Speaker A's second response. Thus, Speaker A's gazes to Speaker B's objects prior to the onset of their second response may have served to glean additional information about Speaker B's task. The short durations of Speaker A's gazes to the partner's objects in the ABA condition suggests that they most likely recognized the objects and initiated, but did not complete access to their names. Representing this kind of information might support the timing of Speaker A's utterances because part of the variability in the time required to name objects arises during the early stages of conceptual and speech planning and because uncommon objects, which are difficult to identify, tend to have names that are difficult to retrieve (e.g., Alario, Ferrand, Laganaro, New, Frauenfelder, & Segui, 2004; Severens et al., 2005; Stadthagen-Gonzalez, Damian, Pérez, Bowers, & Marín, 2009). Thus, knowing that the partner was planning to name an object that was easy or harder to recognize or name may have informed Speaker A's decision when to initiate their own utterance.

If this suggestion is correct, the coordination of the speakers' utterances should be tighter on trials on which they fixated upon the pictures named by their partner than on trials on which they did not do so. However, an analysis of the pauses between the second and third object name in the ABA condition of Experiment 2 showed the opposite pattern: Pauses were slightly more frequent (occurring on 92% vs. 85% of the trials) and longer (258 ms vs. 224 ms) when Speaker A had inspected the second object than when they had not done so. Though the results of this post hoc analysis should not be overinterpreted, they do not support the suggestion that the early gazes to B's object facilitated A's utterance timing. Instead these gazes might also primarily have served to monitor the accuracy of the utterance: Speaker A usually inspected B's object after they had completed their own utterance, but occasionally they did so when Speaker B was about to name the object. In Experiment 1, this latter strategy may have been more strongly encouraged than in Experiment 2 as Speaker B's object was located between the two objects to be named by Speaker A.

We observed that in both experiments, good temporal coordination of the turns was achieved without evidence for full corepresentation of the partner's task share. As already indicated, some representation of the partner's task had to be present to follow the task instructions to name some, but not all of the objects and, in the ABA condition, allow the partner to complete the turn. Timely resuming the turn after Speaker B's utterance was likely facilitated by the fact that both participants knew how many pictures each of them had to name and that they knew from the familiarization phase and earlier trials that the pictures were quite similar in character. In the ABA condition, Participant A knew that their partner would produce exactly one object name between their own two object names. Thus, they could build upon their own experience or their memory of the partner's performance on earlier trials to predict how much time their partner would need to name a picture on a given trial. In other words, as in the study by Gambi and colleagues, speakers might not have represented which specific picture their partner was naming, but simply that the partner was naming one picture.

Another strategy the speakers could use to time their utterances would be to prepare an object name as early as possible and launch articulation contingent upon the partner's speech. We observed

average gap durations around 250 ms between the speakers' turns. The estimated minimum reaction time (RT) for a prepared spoken response is around 200 ms (e.g., Fry, 1975), but in many delayed naming experiments latencies of at least 400 ms have been observed (e.g., Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Barry, Hirsh, Johnston, & Williams, 2001; Piai, Roelofs, & Schriefers, 2011). In the present experiments, the average spoken word durations were around 500 ms. Therefore, the participants in the present experiments could have launched their planned turns shortly after they had heard the onset of the object name produced by their partner and still respond within 200 ms of the offset of the partner's turn.

Post hoc analysis showed that in Experiment 1 the duration of the silence between Speaker A's naming of the first object and Speaker B's naming of the second object (collapsing across the ABx and ABA conditions) was longer when Speaker A's object had a monosyllabic name ($M = 291$ ms; $N = 55$ objects) than when it had a disyllabic name ($M = 265$ ms; $N = 22$ objects; three objects with trisyllabic names were not included in the analysis; $B = -.14$, $SE = .07$, $t = -2.11$). In Experiment 2, this effect was present numerically (pause following monosyllabic names: $M = 235$ ms; pause following disyllabic names: $M = 218$ ms), but did not reach statistical significance ($B = -.17$, $SE = .11$, $t = -1.48$). For the second interturn pause in the ABA condition, the same nonsignificant trend was observed in Experiment 1 (monosyllabic names: $M = 279$; $N = 40$ objects, disyllabic names: $M = 250$; $N = 31$ objects; $t < 1$), but not in Experiment 2 (monosyllabic names: $M = 236$; disyllabic names: $M = 245$; $t < 1$). The numerical trend of shorter pauses following longer words indicates that launching articulation upon hearing the onset of a partner's turn may have been one of the strategies speakers used to coordinate timing in this task. This strategy does not require corepresentation, as instead of representing the partner's speech plan, participants could rely on the partner's actual speech to guide the timing of their response. Note, however, that this pattern of results is also consistent with the proposal that speakers based the timing of their utterances on an average of the partner's word durations gleaned from earlier trials; when such an average is used pauses after short words would be expected to be slightly shorter than pauses after longer words.

In sum, speakers in the present study most likely achieved the coordination of their utterances by relying on partial corepresentation of the partner's task and using the actual speech produced by their partner. How systematically corepresentation occurs in other contexts and how well it supports speakers' attainment of different goals, such as utterance timing and monitoring carefully what is being said, needs to be determined in further work. This could be done in experiments varying how much information speakers have about their partners' task (e.g., whether or not they can see the partner's objects or know the number of objects the partner has to name) how well or predictably the partner carries out the task and how important exact timing of the utterances or monitoring of their correctness is. It would also be important to determine how specific characteristics of the procedure used in the current experiment affected the participants' coordination strategies. For instance, the use of pictures to elicit the utterances should encourage corepresentation, as the speakers knew precisely what to represent. The use of the picture frames to indicate each speaker's task may have affected their allocation of visual attention and may have

biased them toward or against corepresentation. The impact of these and other features of the experimental procedure can be assessed in further work.

An important more general question than how temporal coordination is achieved in various experimental paradigms is how it is achieved in everyday dialogue. The two-object joint naming (ABx) condition of the present study is structurally similar to a question-answer pair, and the three-object naming condition (ABA) resembles a question-answer exchange with an introductory question (e.g., Speaker A: "Do you serve decaf?" Speaker B: "We do." Speaker A: "Two, please."). In addition, these structures have been likened to collaborative turn completions (Gambi & Pickering, 2011; Lerner, 1991). Given that participants in the present experiments appeared to draw on several strategies to achieve coordination, it seems like that speakers in such everyday exchanges do the same. Sometimes they fully corepresent their partner's utterance and can then complete it using the exact same words the partner would have chosen themselves. Perhaps more commonly, speakers generate less well-defined notions of their partner's upcoming speech and use them to prepare their reply, or they wait until the other person has completed their turn and then respond.

This view is consistent with the framework for conversational turn-taking developed by Levinson and Torreira (2015). They proposed that interlocutors in natural conversations can often guess their partner's speech act (e.g., whether it is a request or a statement) and the gist of the utterance early during a turn and immediately use this information to begin to plan their response. Listeners are sensitive to cues predicting the end of the current turn and launch their prepared utterance shortly before the predicted end of turn. This view entails that interlocutors pay attention to each other's utterances (as indeed one would hope) and predict more or less precisely what their partner might say next, but it does not involve corepresentation to the level of individual words.

In considering coordination in conversation it is important to keep in mind that this coordination is far from perfect. As has often been stressed, gaps between turns tend to be short, with modes under 300 ms (e.g., Stivers et al., 2009). Yet, there is substantial variation around these modes. Analyzing three corpora of conversational speech in three different languages, Heldner and Edlund (2010) observed that the duration of between-speaker pauses in the corpora was highly distributed, with 18% to 30% of turn-transition intervals exceeding 500 ms. Though unexpectedly long pauses carry meaning (for instance, expressing reluctance to agree, Bögels, Kendrick, & Levinson, 2015; Kendrick & Torreira, 2015), and frequent mutual interruptions may disrupt comprehension and be perceived as impolite, overall interlocutors appear not to be overly sensitive to some variation in the duration of the gaps between their turns. As Heldner and Edlund pointed out, many between-speaker silent intervals are long enough for the next turn to be a response to the end of the interlocutor's turn, rather than a projection of the upcoming turn end. This detection-based (as opposed to projection-based) strategy would lead to sufficient coordination in the present study and might be sufficient in many other situations as well.

It is also worth remembering that in everyday conversations speakers can make use of a diverse array of social cues to mitigate the possible consequences of imprecise turn timing. For instance, speakers may use fillers such as "ehm" or "hm," or nonverbal reactions such as nods and bodily shifts to signal engagement to

their interlocutor and indeed some corpus analyses of conversational interaction have classified nonverbal reactions as responses in the same way as verbal utterances (e.g., Stivers et al., 2009). This ability to respond or signal engagement with minimal verbal utterances or nonverbally relaxes the constraints on the timing of verbal utterances, reducing the necessity of corepresentation as a mechanism for coordination.

Conversation is often likened to joint action, such as carrying a couch up a flight of stairs. In their review of the mechanisms underlying joint action, Vesper and colleagues (2016, 2017) stressed that corepresentation of the partner's task is one of several mechanisms people draw upon when they accomplish joint tasks. Other mechanisms include paying attention to the physical environment (e.g., noticing the bend in the stairs), remembering how the task was carried out on earlier occasions, and reacting to the partner's actions (e.g., by holding on to the couch more forcefully when the partner has left). Moreover, only some joint actions (such as playing in an orchestra) require perfect synchronization; in many everyday task, including carrying furniture, the partners may only aim for and achieve rough coordination of their actions. The same may be true for most types of everyday conversation: Millisecond accuracy of utterance timing is not required.

To conclude, the current study demonstrated that interlocutors in a shared naming task were affected by an awareness of each other's task, but found no evidence that speakers represented each other's naming assignment as if it were part of their own speech task. This does not rule out that full corepresentation occurs in other situations, but it does provide evidence that corepresentation is not required for smooth turn taking and suggests that speakers may not corepresent when other strategies for coordination are available. In everyday conversation, sufficient coordination is probably achieved through multiple means, including partial corepresentation and simply responding to the preceding turn.

References

- Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*, 140–155. <http://dx.doi.org/10.3758/BF03195559>
- Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review*, *14*, 1177–1182. <http://dx.doi.org/10.3758/BF03193109>
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264. [http://dx.doi.org/10.1016/S0010-0277\(99\)00059-1](http://dx.doi.org/10.1016/S0010-0277(99)00059-1)
- Aristei, S., Zwitserlood, P., & Abdel Rahman, R. (2012). Picture-induced semantic interference reflects lexical competition during object naming. *Frontiers in Psychology*, *3*, 28. <http://dx.doi.org/10.3389/fpsyg.2012.00028>
- Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research*, *211*(3–4), 371–385. <http://dx.doi.org/10.1007/s00221-011-2709-9>
- Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action corepresentation: The joint SNARC effect. *Social Neuroscience*, *3*(3–4), 410–420. <http://dx.doi.org/10.1080/17470910801900908>
- Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of

- picture naming. *Journal of Memory and Language*, 44, 350–375. <http://dx.doi.org/10.1006/jmla.2000.2743>
- 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, 1858. <http://dx.doi.org/10.3389/fpsyg.2016.01858>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and S4 (R Package, Version 1[7]). *Journal of Statistical software*. Los Angeles, CA: Department of Statistics, University of California, Los Angeles.
- Baus, C., Sebanz, N., de la Fuente, V., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition*, 133, 395–407. <http://dx.doi.org/10.1016/j.cognition.2014.07.006>
- Bekkering, H., de Bruijn, E. R., Cuijpers, R. H., Newman-Norlund, R., Van Schie, H. T., & Meulenbroek, R. (2009). Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science*, 1, 340–352. <http://dx.doi.org/10.1111/j.1756-8765.2009.01023.x>
- Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 1404–1415. <http://dx.doi.org/10.1037/a0027523>
- Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer*. [Computer program]. Retrieved from <http://www.praat.org/>
- Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). *Never say no: How the brain interprets the pregnant pause in conversation*. Retrieved from <http://journals.plos.org/plosone/article?id=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. <http://dx.doi.org/10.1038/srep12881>
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, 53, 301–344. [http://dx.doi.org/10.1016/S0079-7421\(10\)53008-1](http://dx.doi.org/10.1016/S0079-7421(10)53008-1)
- Clark, H. H. (1996). *Using language*. New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511620539>
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12, 453–459. <http://dx.doi.org/10.3758/BF03193787>
- Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, 10, 111–117. <http://dx.doi.org/10.3758/BF03196474>
- Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369, 20120394. <http://dx.doi.org/10.1098/rstb.2012.0394>
- Demiral, Ş. B., Gambi, C., Nieuwland, M. S., & Pickering, M. J. (2016). Neural correlates of verbal joint action: ERPs reveal common perception and action systems in a shared-stroop task. *Brain Research*, 1649 Part A, 79–89.
- De Ruiter, J., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82, 515–535. <http://dx.doi.org/10.1353/lan.2006.0130>
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 491–505. <http://dx.doi.org/10.1111/j.1469-8986.2007.00531.x>
- Freedman, M. L., Martin, R. C., & Biegler, K. (2004). Semantic relatedness effects in conjoined noun phrase production: Implications for the role of short-term memory. *Cognitive Neuropsychology*, 21, 245–265. <http://dx.doi.org/10.1080/02643290342000528>
- Fry, D. B. (1975). Simple reaction-times to speech and non-speech stimuli. *Cortex*, 11, 355–360. [http://dx.doi.org/10.1016/S0010-9452\(75\)80027-X](http://dx.doi.org/10.1016/S0010-9452(75)80027-X)
- Gambi, C., Cop, U., & Pickering, M. J. (2015a). How do speakers coordinate? Evidence for prediction in a joint word-replacement task. *Cortex*, 68, 111–128. <http://dx.doi.org/10.1016/j.cortex.2014.09.009>
- Gambi, C., & Pickering, M. J. (2011). A cognitive architecture for the coordination of utterances. *Frontiers in Psychology*, 2, 275. <http://dx.doi.org/10.3389/fpsyg.2011.00275>
- Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015b). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 1–21. <http://dx.doi.org/10.1037/a0037438>
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181–218. [http://dx.doi.org/10.1016/0010-0277\(87\)90018-7](http://dx.doi.org/10.1016/0010-0277(87)90018-7)
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1, 292–304. <http://dx.doi.org/10.1111/j.1756-8765.2009.01020.x>
- Gordon, P. C., & Hoedemaker, R. S. (2016). Effective scheduling of looking and talking during rapid automatized naming. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 742–760. <http://dx.doi.org/10.1037/xhp0000171>
- Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82(1), B1–B14. [http://dx.doi.org/10.1016/S0010-0277\(01\)00138-X](http://dx.doi.org/10.1016/S0010-0277(01)00138-X)
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279. <http://dx.doi.org/10.1111/1467-9280.00255>
- Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. *Cognition*, 108, 601–607. <http://dx.doi.org/10.1016/j.cognition.2008.04.005>
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38, 555–568. <http://dx.doi.org/10.1016/j.wocn.2010.08.002>
- 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. <http://dx.doi.org/10.1016/j.actpsy.2016.11.007>
- Holländer, A., Jung, C., & Prinz, W. (2011). Covert motor activity on NoGo trials in a task sharing paradigm: Evidence from the lateralized readiness potential. *Experimental Brain Research*, 211(3–4), 345–356. <http://dx.doi.org/10.1007/s00221-011-2688-x>
- Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research PRPF*, 73, 512–526. <http://dx.doi.org/10.1007/s00426-009-0234-2>
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). Codes and their vicissitudes. *Behavioral and Brain Sciences*, 24, 910–926. <http://dx.doi.org/10.1017/S0140525X01520105>
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, 100, 464–482. <http://dx.doi.org/10.1016/j.cognition.2005.02.006>
- Huetig, F., & Altmann, G. T. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, 15, 985–1018. <http://dx.doi.org/10.1080/13506280601130875>
- Huetig, F., & Hartsuiker, R. J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition*, 36, 341–360. <http://dx.doi.org/10.3758/MC.36.2.341>
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144. <http://dx.doi.org/10.1016/j.cognition.2002.06.001>

- Jescheniak, J. D., Matushanskaya, A., Mädebach, A., & Müller, M. M. (2014). Semantic interference from distractor pictures in single-picture naming: Evidence for competitive lexical selection. *Psychonomic Bulletin & Review*, *21*, 1294–1300. <http://dx.doi.org/10.3758/s13423-014-0606-5>
- Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of processing in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156. [http://dx.doi.org/10.1016/S0749-596X\(03\)00023-8](http://dx.doi.org/10.1016/S0749-596X(03)00023-8)
- Kapoula, Z., & Robinson, D. A. (1986). Saccadic undershoot is not inevitable: Saccades can be accurate. *Vision Research*, *26*, 735–743. [http://dx.doi.org/10.1016/0042-6989\(86\)90087-8](http://dx.doi.org/10.1016/0042-6989(86)90087-8)
- Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*, 255–289. <http://dx.doi.org/10.1080/0163853X.2014.955997>
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. <http://dx.doi.org/10.3758/BRM.42.3.643>
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In R. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 59–101). New York, NY: Academic Press. <http://dx.doi.org/10.1016/B978-0-12-385527-5.00003-6>
- 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. <http://dx.doi.org/10.1016/j.cognition.2017.05.024>
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59. <http://dx.doi.org/10.1080/23273798.2015.1102299>
- Lerner, G. H. (1991). On the syntax of sentences-in-progress. *Language in Society*, *20*, 441–458. <http://dx.doi.org/10.1017/S0047404500016572>
- Levelt, W. J., & Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. *The European Journal of Cognitive Psychology*, *12*, 433–452. <http://dx.doi.org/10.1080/095414400750050178>
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731. <http://dx.doi.org/10.3389/fpsyg.2015.00731>
- Loftus, G. F., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.
- 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. <http://dx.doi.org/10.1037/a0018522>
- Matushanskaya, A., Mädebach, A., Müller, M. M., & Jescheniak, J. D. (2016). When sufficiently processed, semantically related distractor pictures hamper picture naming. *Experimental Psychology*, *63*, 307–317. <http://dx.doi.org/10.1027/1618-3169/a000340>
- McConkie, G. W., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Attention, Perception & Psychophysics*, *46*, 245–253. <http://dx.doi.org/10.3758/BF03208086>
- Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*, 477–496. <http://dx.doi.org/10.1006/jmla.1996.0026>
- Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language*, *57*, 210–231. <http://dx.doi.org/10.1016/j.jml.2006.10.005>
- Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, *35*, 494–503. <http://dx.doi.org/10.3758/BF03193289>
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33. [http://dx.doi.org/10.1016/S0010-0277\(98\)00009-2](http://dx.doi.org/10.1016/S0010-0277(98)00009-2)
- Meyer, A. S., Wheeldon, L., Van der Meulen, F., & Konopka, A. (2012). Effects of speech rate and practice on the allocation of visual attention in multiple object naming. *Frontiers in psychology*, *3*, 39. <http://dx.doi.org/10.3389/fpsyg.2012.00039>
- Morgan, J. L., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 428–442. <http://dx.doi.org/10.1037/0278-7393.31.3.428>
- Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*, 359–377. <http://dx.doi.org/10.1016/j.jml.2005.05.001>
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227–252. <http://dx.doi.org/10.1016/j.cognition.2009.09.007>
- 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. <http://dx.doi.org/10.1016/j.jml.2011.01.004>
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*, 105–110. <http://dx.doi.org/10.1016/j.tics.2006.12.002>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 329–347. <http://dx.doi.org/10.1017/S0140525X12001495>
- Powell, M. J. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. (Tech. Rep. No. DAMTP NA2009/06). Cambridge, UK: Cambridge University, Department of Applied Mathematics and Theoretical Physics.
- Prinz, W. (1997). Perception and action planning. *The European Journal of Cognitive Psychology*, *9*, 129–154. <http://dx.doi.org/10.1080/713752551>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. <http://dx.doi.org/10.1037/0033-2909.124.3.372>
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, *18*, 407–413. <http://dx.doi.org/10.1111/j.1467-9280.2007.01914.x>
- Richardson, D. C., Street, C. N. H., Tan, J. Y. M., Kirkham, N. Z., Hoover, M. A., & Ghane Cavanaugh, A. (2012). Joint perception: Gaze and social context. *Frontiers in Human Neuroscience*, *6*, 194. <http://dx.doi.org/10.3389/fnhum.2012.00194>
- Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organisation on the timing of turn taking: A corpus study. *Frontiers in psychology*, *6*, 509.
- Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language*, *57*, 232–251. <http://dx.doi.org/10.1016/j.jml.2006.10.001>
- Roelofs, A. (2008). Tracing attention and the activation flow of spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 353–368. <http://dx.doi.org/10.1037/0278-7393.34.2.353>
- 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. <http://dx.doi.org/10.1037/a0028646>
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70–76. <http://dx.doi.org/10.1016/j.tics.2005.12.009>

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science, 1*, 353–367. <http://dx.doi.org/10.1111/j.1756-8765.2009.01024.x>

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), B11–B21. [http://dx.doi.org/10.1016/S0010-0277\(03\)00043-X](http://dx.doi.org/10.1016/S0010-0277(03)00043-X)

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. <http://dx.doi.org/10.1016/j.actpsy.2005.01.002>

Stadthagen-Gonzalez, H., Damian, M. F., Pérez, M. A., Bowers, J. S., & Marín, J. (2009). Name-picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62*, 1581–1597. <http://dx.doi.org/10.1080/17470210802511139>

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. <http://dx.doi.org/10.1073/pnas.0903616106>

Team, R. C. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Tydgat, I., Diependaele, K., Hartsuiker, R. J., & Pickering, M. J. (2012). How lingering representations of abandoned context words affect speech production. *Acta Psychologica, 140*, 218–229. <http://dx.doi.org/10.1016/j.actpsy.2012.02.004>

Tydgat, I., Stevens, M., Hartsuiker, R. J., & Pickering, M. J. (2011). Deciding where to stop speaking. *Journal of Memory and Language, 64*, 359–380. <http://dx.doi.org/10.1016/j.jml.2011.02.002>

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467. <http://dx.doi.org/10.1037/0278-7393.31.3.443>

van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. (2001). Eye movements during the production of nouns and pronouns. *Memory & Cognition, 29*, 512–521. <http://dx.doi.org/10.3758/BF03196402>

Veenstra, A., Meyer, A. S., & Acheson, D. J. (2015). Effects of parallel planning on agreement production. *Acta Psychologica, 162*, 29–39. <http://dx.doi.org/10.1016/j.actpsy.2015.09.011>

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., . . . Wahn, B. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology, 7*, 2039. <http://dx.doi.org/10.3389/fpsyg.2016.02039>

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks, 23*(8–9), 998–1003. <http://dx.doi.org/10.1016/j.neunet.2010.06.002>

Vesper, C., Schmitz, L., Safra, L., Sebanz, N., & Knoblich, G. (2016). The role of shared visual information for joint action coordination. *Cognition, 153*, 118–123. <http://dx.doi.org/10.1016/j.cognition.2016.05.002>

Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011). What is shared in joint action? issues of co-representation, response conflict, and agent identification. *Review of Philosophy and Psychology, 2*, 147–172. <http://dx.doi.org/10.1007/s13164-011-0057-0>

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*, 460–473. <http://dx.doi.org/10.1037/0033-2909.131.3.460>

Appendix

Object Triplets Named in Experiment 1 and 2 (Dutch Labels in Parentheses)

Object 1	Object 2 (related)	Object 2 (unrelated)	Object 3
necklace (ketting)	earring (oorbel)	sink (wasbak)	accordion (accordeon)
bath tub (bad)	sink (wasbak)	earring (oorbel)	thumb (duim)
map (kaart)	globe (wereldbol)	desk (bureau)	belt (riem)
table (tafel)	desk (bureau)	globe (wereldbol)	barrel (ton)
car (auto)	bicycle (fiets)	fence (hek)	cross (kruis)
wall (muur)	fence (hek)	bicycle (fiets)	bottle (fles)
shovel (schop)	ax (bijl)	paw (poot)	clown (clown)
hoof (hoef)	paw (poot)	ax (bijl)	bird (vogel)
duck (eend)	swan (zwaan)	cherry (kers)	branch (tak)
apple (appel)	cherry (kers)	swan (zwaan)	beetle (kever)
balloon (ballon)	kite (vlieger)	zebra (zebra)	cupboard (kast)
horse (paard)	zebra (zebra)	kite (vlieger)	diamond (diamant)
goat (geit)	sheep (schaap)	toe (teen)	doorknob (deurknop)
finger (vinger)	toe (teen)	sheep (schaap)	corn (mais)
refrigerator (koelkast)	stove (fornuis)	brush (borstel)	butterfly (vlinder)
comb (kam)	brush (borstel)	stove (fornuis)	dragon (draak)
guitar (gitaar)	trumpet (trompet)	canoe (kano)	eye (oog)
boat (boot)	canoe (kano)	trumpet (trompet)	fire (vuur)
heart (hart)	lungs (longen)	igloo (iglo)	mountain (berg)
tent (tent)	igloo (iglo)	lungs (longen)	gorilla (gorilla)

(Appendix continues)

Appendix (continued)

Object 1	Object 2 (related)	Object 2 (unrelated)	Object 3
lion (leeuw)	tiger (tijger)	knight (ridder)	needle (naald)
king (koning)	knight (ridder)	tiger (tijger)	neck (nek)
roller skate (rolschaats)	jump rope (springtouw)	cup (kopje)	owl (uil)
glass (glas)	cup (kopje)	jump rope (springtouw)	parachute (parachute)
letter (brief)	package (pakje)	cookie (koekje)	pillow (kussen)
cake (taart)	cookie (koekje)	package (pakje)	ring (ring)
suitcase (koffer)	backpack (rugzak)	cigar (sigaar)	pitcher (kan)
pipe (pijp)	cigar (sigaar)	backpack (rugzak)	present (cadeau)
cheese (kaas)	butter (boter)	ladder (ladder)	raft (vlot)
stairs (trap)	ladder (ladder)	butter (boter)	skeleton (skelet)
pig (varken)	cow (koe)	thread (draad)	submarine (duikboot)
rope (touw)	thread (draad)	cow (koe)	shoulder (schouder)
sun (zon)	moon (maan)	fly (vlieg)	smoke (rook)
bee (bij)	fly (vlieg)	moon (maan)	seal (zeehond)
couch (bank)	chair (stoel)	railroad tracks (spoor)	stool (kruk)
road (weg)	railroad tracks (spoor)	chair (stoel)	sausage (worst)
spider (spin)	ant (mier)	candle (kaars)	trash (vuilnis)
lamp (lamp)	candle (kaars)	ant (mier)	vacuum (stofzuiger)
rabbit (konijn)	squirrel (eekhoorn)	pencil (potlood)	vase (vaas)
pen (pen)	pencil (potlood)	squirrel (eekhoorn)	waiter (ober)
pants (broek)	dress (jurk)	fountain (fontein)	well (put)
statue (beeld)	fountain (fontein)	dress (jurk)	pool (zwembad)
fork (vork)	spoon (lepel)	bus (bus)	wig (pruik)
train (trein)	bus (bus)	spoon (lepel)	wheelchair (rolstoel)
door (deur)	window (raam)	skis (ski's)	spaghetti (spaghetti)
sled (slee)	skis (ski's)	window (raam)	house (huis)
tree (boom)	shrub (struik)	rainbow (regenboog)	chest (kist)
cloud (wolk)	rainbow (regenboog)	shrub (struik)	watering can (gieter)
frog (kikker)	snail (slak)	leg (been)	whistle (fluit)
arm (arm)	leg (been)	snail (slak)	paintbrush (penseel)
bowl (kom)	plate (bord)	parrot (papegaai)	traffic light (stoplicht)
eagle (adelaar)	parrot (papegaai)	plate (bord)	anchor (anker)
dog (hond)	cat (kat)	clock (klok)	witch (heks)
watch (horloge)	clock (klok)	cat (kat)	map (kaart)
shark (haai)	whale (walvis)	butcher (slager)	paper clip (paperclip)
cook (kok)	butcher (slager)	whale (walvis)	wing (vleugel)
broom (bezem)	rake (hark)	carrot (wortel)	rain (regen)
lettuce (sla)	carrot (wortel)	rake (hark)	bow (strik)
microphone (microfoon)	speaker (boks)	cannon (kanon)	potato (aardappel)
gun (pistool)	cannon (kanon)	speaker (boks)	city (stad)
pot (pan)	kettle (ketel)	hat (muts)	bomb (bom)
scarf (sjaal)	hat (muts)	kettle (ketel)	wizard (tovenaar)
nose (neus)	lips (lippen)	castle (kasteel)	rose (roos)
farm (boerderij)	castle (kasteel)	lips (lippen)	alarm clock (wekker)
hammer (hamer)	drill (boor)	blouse (blouse)	coat (jas)
sweater (trui)	blouse (blouse)	drill (boor)	star (ster)
lemon (citroen)	orange (sinaasappel)	radio (radio)	flag (vlag)
TV (tv)	radio (radio)	orange (sinaasappel)	net (net)
hat (hoed)	cap (pet)	ghost (spook)	key (sleutel)
genie (geest)	ghost (spook)	cap (pet)	banana (banaan)
faucet (kraan)	hose (tuinslang)	swing (schommel)	bread (brood)
seesaw (wip)	swing (schommel)	hose (tuinslang)	monkey (aap)
button (knoop)	zipper (rits)	basket (mand)	ball (bal)
box (doos)	basket (mand)	zipper (rits)	bell (bel)
airplane (vliegtuig)	helicopter (helikopter)	binoculars (verrekijker)	rock (steen)
glasses (bril)	binoculars (verrekijker)	helicopter (helikopter)	doll (pop)
bed (bed)	crib (wieg)	sock (sok)	tear (traan)
glove (handschoen)	sock (sok)	crib (wieg)	lock (slot)
hand (hand)	foot (voet)	scissors (schaar)	roof (dak)
knife (mes)	scissors (schaar)	foot (voet)	book (boek)

Note. In Experiment 2, the following low-accuracy objects were replaced (replacement object in bold): eagle (adelaar), **baby carriage (kinderwagen)**; bee (bij), **feather (veer)**; farm (boerderij), **crab (krab)**; suitcase (koffer), **ear (oor)**; sink (wasbak), **bone (bot)**; crib (wieg), **slide (glijbaan)**.

Received July 7, 2017

Revision received December 29, 2017

Accepted March 2, 2018 ■