

Mental representations of partner task cause interference in picture naming[☆]

Laurel Brehm^{a,*}, Linda Taschenberger^b, Antje Meyer^a

^a Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

^b University College, London, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Keywords:

Language production
Picture naming
Interference
Attention
Language as joint action

ABSTRACT

Interference in picture naming occurs from representing a partner's preparations to speak (Gambi, van de Cavey, & Pickering, 2015). We tested the origins of this interference using a simple non-communicative joint naming task based on Gambi et al. (2015), where response latencies indexed interference from partner task and partner speech content, and eye fixations to partner objects indexed overt attention. Experiment 1 contrasted a partner-present condition with a control partner-absent condition to establish the role of the partner in eliciting interference. For latencies, we observed interference from the partner's task and speech content, with interference increasing due to partner task in the partner-present condition. Eye-tracking measures showed that interference in naming was not due to overt attention to partner stimuli but to broad expectations about likely utterances. Experiment 2 examined whether an equivalent non-verbal task also elicited interference, as predicted from a language as joint action framework. We replicated the finding of interference due to partner task and again found no relationship between overt attention and interference. These results support Gambi et al. (2015). Individuals co-represent a partner's task while speaking, and doing so does not require overt attention to partner stimuli.

1. Introduction

Conversation has been seen as a form of joint action: it involves two or more people coordinating their actions—speaking, gesturing, and other types of non-verbal behavior—to reach the common goal of understanding each other. Research on joint motor action has shown that in order for people to act jointly, each agent must represent and predict the other's actions (see e.g. Knoblich, Butterfill, & Sebanz, 2011). Many authors have argued that the same is true for conversation (e.g., Clark, 1996; Pickering & Garrod, 2004, 2013). Representation of the partner's upcoming utterances is, for instance, often seen as a prerequisite of smooth turn-taking (e.g., Levinson, 2016; Levinson & Torreira, 2015). An important question arising from this view of conversation concerns the nature of the representations that interlocutors generate of their partners' utterance plans. Too little representation of the partner's utterance plan could lead to insufficient coordination to converse smoothly, while too much representation may be costly and lead to interference in speech planning. In the current experiments, we tracked the allocation of speakers' overt visual attention to objects named by a

partner and measured interference in speech planning to assess how the speakers represented their interlocutors' utterances. The overarching goal was to provide insight into the mechanisms that support conversation.

Earlier work by Gambi, van de Cavey, and Pickering (2015) showed that speakers performing a joint naming task represented the partner's task, i.e. whether or not the partner also had to name objects, but not the content of the partner's speech plan. This suggests that what is ideal for conversation may involve only minimally representing a partner's plan. The current work followed directly from this study, using a simplified version of the Gambi et al. (2015) paradigm. As such, we begin by describing their procedure and results in detail.

In the Gambi et al. (2015) paradigm, two participants were seated in separate rooms and asked to name colored superimposed pictures in a particular order (e.g. a large *apple* in red, with a smaller *blouse* in blue nested inside, eliciting the response *apple, blouse*). Participants were cued with a screen of instructions before each trial as to whether their partner was either naming the pictures (in the same order or the opposite order) or performing a different task (withholding a response or

[☆] Author note: Portions of this work were included in Linda Taschenberger's MA thesis at the University of Bonn. We would like to thank Marwa Mekni Toujani for assistance with Experiment 2. Data and scripts are archived on the MPI language archive: <https://tla.mpi.nl>.

* Corresponding author at: Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, the Netherlands.

E-mail address: laurel.brehm@mpi.nl (L. Brehm).

deciding whether the two pictures belonged to the same semantic category). The goal of the experiments was to examine whether individuals only represented that the partner was or was not also engaged in a naming task, or whether they represented the content of the partner's utterances as well.

Gambi and colleagues' Experiment 1 and Experiment 2 examined how the average onset latency and accuracy of two-word utterances were affected by the task a partner was simultaneously performing. In Experiment 1, the authors contrasted the degree of interference in naming that occurred when a partner was naming the same two pictures in the same order, naming the same two pictures in a different order, or producing no response. Individuals took longer to name their own pictures when they believed their partner was also naming, versus withholding a response. For the naming latencies, there was no effect of the content of the partner's utterance, with no difference in latencies between the same-order or different-order conditions. However, individuals did produce more errors when they believed their partner was preparing a different response. This suggests that individuals received more interference in naming from representing a partner's task than the content of their speech. Experiment 2 replaced the no-response condition with a categorization task in which the individual had to state whether the two pictures belonged to the same semantic category using a verbal yes/no response. Again, naming latencies were slower when the partner was believed to be naming than categorizing. However, there was no effect of the content of the partner's utterance (same-picture versus different-picture condition) on either naming latencies or error rates. This provides further support for the claim that individuals derived more interference from representing the partner's task than the content of their speech.

Gambi and colleagues' Experiment 3 and Experiment 4 examined how the onset latency and accuracy of one-word utterances were affected by the partner's task. Experiment 3 used identical pictures to Experiments 1 and 2 but required the participant to name only one of them, contrasting a same-picture condition with different-picture and no-response conditions. This experiment showed no evidence for interference from either partner task or partner utterance on naming latencies or error rates. Experiment 4 used the same design as Experiment 3 but visually degraded the pictures to make naming more difficult. This replicated Experiment 1: naming latencies were slower when the partner was believed to also be naming than not responding. More errors were again produced when the partner was believed to be producing a different response, though only when the partner's picture was the larger of the two nested objects.

In sum, the results from Gambi et al. (2015) clearly demonstrated that participants represented whether or not their partner was speaking. In addition, error patterns and the latencies in Experiment 4 suggested that the content of partner utterances may also be occasionally or weakly represented. This suggests that individuals formed a partial co-representation of the partner's utterance across multiple levels, representing the fact that speech was occurring and also representing a sparse amount of information about the speech content.

The fact that naming interference was observed at all is intriguing given that the task in Gambi et al. (2015) was non-communicative, with no joint task goals, and given that the two interlocutors sat in different rooms. This means that there was no direct benefit to representing the partner's speech plan, and participants could have in fact completely ignored the partner without any consequences for their task performance. One possible explanation for these results is therefore that co-representation arises from fundamental, automatic processes activated when another person is perceived as important for participating in a joint task, even if they are not present in the same room (see e.g. Kuhlén & Rahman, 2017, for supporting results). Another possible explanation is that the specific pattern of co-representation arose due to specific parameters of the paradigm, such as the trial-specific instructions (serving to provide a clear reminder of the participant and partner's tasks before each trial) or the stimulus design (presenting two nested

objects, making it hard to ignore the partner items). Our experiments were designed to follow up on these two possibilities, examining the role of co-presence on interference in naming in a simplified paradigm. In particular, we wanted to establish whether task and/or content interference appears in a simplified, less demanding version of the paradigm where the partner was physically co-present. We also measured the amount of overt visual attention that partner objects received by recording the participants' eye movements. This allowed us to examine what individuals represent about a partner's preparations to speak, disentangling whether co-representations can be formed from expectations about what is likely given the task paradigm versus from overt visual attention to objects.

1.1. Current study

The goal of the current studies was to examine interference in a simplified version of the joint naming task developed by Gambi et al. (2015). Following earlier work (e.g. Gambi et al., 2015; Hoedemaker, Ernst, Meyer, & Belke, 2017; Hoedemaker & Meyer, 2019; Kuhlén & Rahman, 2017), we presented individuals with pictures to name or categorize and varied properties of the partner stimulus, partner task, and whether there was a partner present during the experimental session. We based the paradigm on Gambi et al. (2015), making changes to make the social context more obvious, simplify the paradigm, change the categorization task, and to be able to make direct observations of attention to partner objects during the experiment.

The first change made from Gambi et al. (2015) was to make the social character of the study more obvious. Rather than being seated in separate rooms, individuals were seated at adjacent computers at the same table and performed the joint naming task in a *partner-present* condition. Each participant wore noise-cancelling headphones so that they could not hear their partner. This served to make the other participant's task and intentions fully transparent while removing interference due to hearing their speech. This means that individuals did not have to infer a partner's co-presence, but could simply observe it. In Experiment 1, this partner-present condition was compared with a *partner-absent* condition in which the participant was told their partner was unable to attend the experiment. This condition provided a baseline with the same visual and task-level complexity, but without any social co-presence.

As in the Gambi et al. (2015) experiments, we asked our participants to perform naming and categorization tasks on identical and different pictures. However, we blocked these conditions rather than interleaving trials of different types. Each block was preceded by instructions that stated what the task would be for each participant (naming or categorizing) and whether the two pictures would be identical or different. This made it so that instructions did not need to appear before each trial, creating a faster and more natural flow. In particular, when both participants were naming, participants might easily conceptualize their task as something analogous to reading aloud together.

In our experiments, we elicited utterances requiring attention to only one object, like in Experiments 3 and 4 of Gambi et al. (2015). Our naming task required naming only one picture, and our categorization task required participants to determine the category of only their picture. This was done to further simplify the paradigm and to allow participants to ignore the partner object if desired, providing a strong test for the automaticity of partner co-representations when not required to perform the task.

The categorization task we used was to decide whether the object was living or not living. In Experiment 1, the judgement was provided using a verbal response, while in Experiment 2, the judgement was provided using a forced-choice button press. This judgement required an identical decision on the part of the participant, but required a simple motor action instead of speaking (see e.g., Knoblich et al., 2011 for an in-depth review of interference in motor action). We varied the output modality in order to disentangle the source of action co-

representations, asking whether individuals specifically represented that their partner is speaking, versus performing another type of action.

Another difference from Gambi et al. (2015) was that we physically separated the two objects, presenting one object on the left side of the screen for one participant and one on the right side for the other participant. The participant seated on the left was tasked with naming or categorizing the left picture, while the participant seated on the right was tasked with naming or categorizing the right picture. This meant that it was always obvious what each participant should say, but the partner object was also always truly irrelevant to the participant's task and easy to ignore if the participant desired to do so.

Separating the two objects also allowed us to collect an additional dependent measure of interference that provided further insight into what sources of information were used to form partner co-representations. We recorded individual's fixations to partner objects in order to track allocation of overt visual attention to competitors during processing (e.g. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; see Hoedemaker & Meyer, 2019, for use of eye-tracking in a joint naming task). Eye-tracking allowed us to test whether overt attention to the partner object was associated with interference in naming. If attention is directly related to interference in naming, the same patterns should be observed on naming latencies and odds of fixating the partner object. Specifically, interference should occur most strongly for different-picture trials when individuals fixated their partner's picture prior to initiating their own response, compared to trials where there was no fixation or only post-speech onset fixations to the partner picture. Collecting this additional dependent measure therefore provided insight into the source and strength of representations activated during naming.

To summarize, Experiment 1 contrasted joint naming in a partner-present condition, in which two participants performed their task side-by-side, with a partner-absent condition, in which one participant performed the task alone. In both conditions, participants named or categorized one picture from a two-picture display, and in the display were either two copies of the same picture or two different pictures. These manipulations were blocked, so that on a given block of trials, all picture pairs were either identical or different for the two participants, and participants either both named or one participant named while the other categorized the pictures.

The goal of Experiment 1 was to assess whether individuals were more susceptible to interference in naming from the non-target object in the partner-present than in the partner-absent condition, which would indicate representation of the partner's task in a social situation (e.g., as in the joint action framework outlined by Knoblich et al., 2011; Sebanz, Bekkering, & Knoblich, 2006; Vesper et al., 2017, and discussed by Gambi et al., 2015). Specifically, we predicted that co-representation of partner task would lead to an interaction between partner presence (present/absent) and partner task (naming/categorizing), with a stronger effect of partner task in the partner-present condition. Likewise, we predicted that co-representation of the partner's task content would lead to an interaction between partner presence and picture congruence (same versus different pictures), with a stronger effect of picture congruence in the partner-present condition. Finally, if interference in naming latencies is caused by active deployment of visual attention to partner pictures, we also predicted that frequent gazes to the partner's picture should be associated with strong partner task and picture congruence effects.

We found that Experiment 1 yielded longer latencies in the partner-present different task than in the partner-present same-task condition, as predicted by a language as joint action framework. The goal of Experiment 2 was to replicate this pattern while requiring a manual response in the categorization task rather than a verbal one. We predicted no effect of partner task if the actions that drive interference are specific to verbal responses, such that only minimal interference should be observed for partner-categorize trials relative to the naming trials when the response is non-verbal. In contrast, we predicted a main effect

of partner task if what matters is purely that the partner is acting, such that robust interference in naming latencies would be observed in partner-categorize trials even when the partner is responding non-verbally.

2. Experiment 1

2.1. Introduction

In the first experiment, participants were assigned to one of two conditions: a partner-present condition, in which two participants performed the same non-communicative naming task in parallel while wearing noise-cancelling headphones, and a partner-absent condition, in which the setup was identical (including two computers displaying stimuli) but only one participant performed the task. Within these between-participants conditions, we manipulated the tasks of the participant and their partner (naming, categorizing), and the congruence of the two stimuli (same, different). If participants represent only their partner's actions, we expect interference based upon partner task but not picture congruence; if participants also represent the content of their partner's utterance, we also expect interference based upon picture congruence.

2.2. Method

2.2.1. Participants

A sample size of 96 participants was selected based upon an a-priori power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), which suggested that 48 participants in each of the between-participant groups would allow observation of an effect f of 0.1 or larger (a small effect size in an ANOVA-style design). 102 participants between the ages of 18 and 30 years (mean age 22 years) were recruited from the Max Planck Institute for Psycholinguistics's database. Six participants were excluded due to calibration issues, track loss, or audio recording failure.

The final sample of 96 participants (71 female) reported normal or corrected-to-normal vision and no speech or hearing problems, and received a payment of 6 € for their participation. Informed consent was obtained from each participant. Of the 96 participants, half were assigned to the partner-absent condition and half to the partner-present condition. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

2.2.2. Materials

Experimental stimuli consisted of thirty black and white line drawings of objects, normed in Belgian Dutch by Severens, van Lommel, Ratnckx, and Hartsuiker (2005), sized to 200 × 200 pixels (3.39° visual angle). Of these, 15 were 'living' (in Dutch: *levend*; five tokens each of animals, fruit, and vegetables) and 15 were non-living (in Dutch: *niet levend*; five tokens each of clothing, furniture, and vehicles). Object names were mono-, di- or tri-syllabic, with the number of syllables not differing between categories ($p = 0.15$). See Appendix A for list of stimuli.

Living and non-living items were matched between sets (all p values above 0.07) based on the Severens et al. (2005) norms for name agreement (mean 91%, range 54% to 100%), log word frequency (mean 1.48, range 0.48 to 3.58), and mean naming time (mean 925 ms, range 694 ms to 1252 ms), as well as on visual complexity, as determined by JPEG file size (mean 15.25, range 5.8 to 24.5). Objects were also assessed for name agreement by six native speakers of Dutch from the Netherlands. This assessment showed that one object from Severens et al. (2005) had a different preferred name in standard Dutch (*stoel* instead of *sessel*); we had participants use this name instead.

2.2.3. Design

Images were presented on two computer monitors with identical

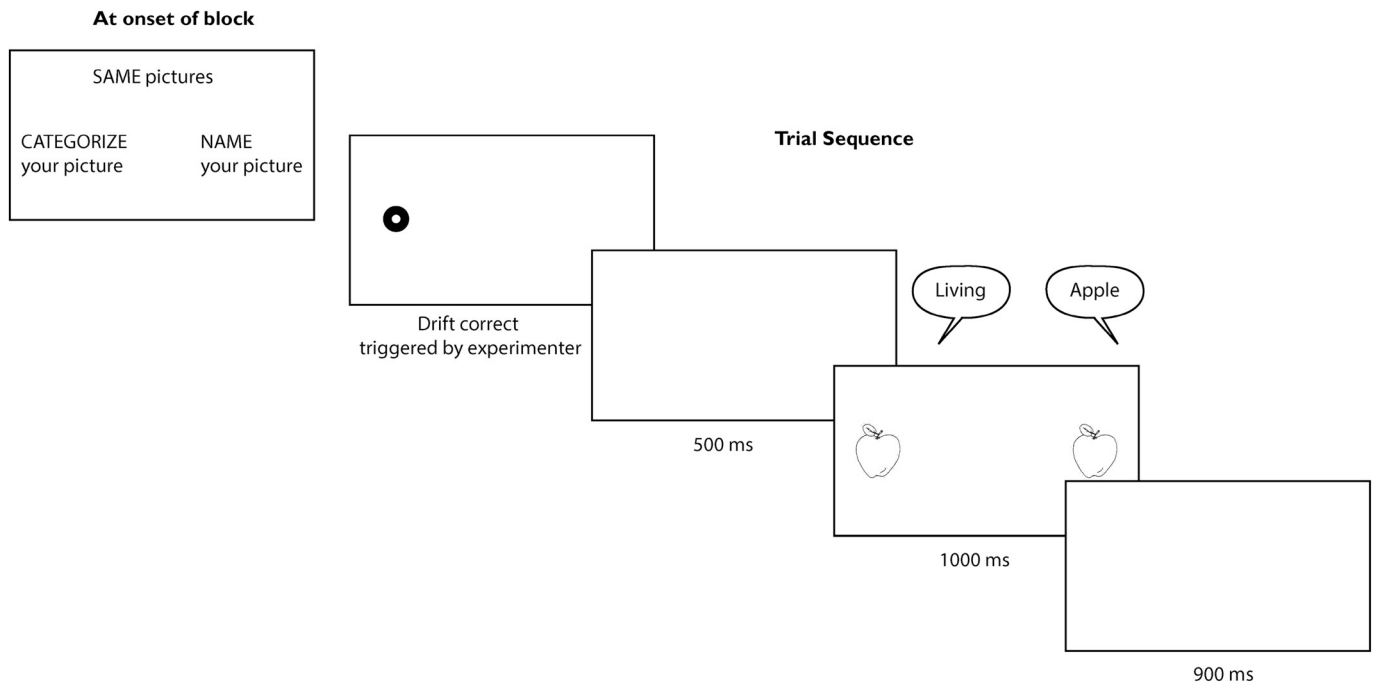


Fig. 1. Sample displays and trial sequence. Fixation point appears on opposite side for other participant.

displays, as shown in Fig. 1. The left object was designated for the participant on the left, and the right object designated for the participant on the right, but both participants were free to view their partner's object on their own monitor. Items were presented in six blocks of 30 trials each.

At the beginning of each block, participants were given a screen of instructions informing them of their task and their partner's task. As in Fig. 1, the instructions *NAME your picture* or *CATEGORIZE your picture* (in Dutch: *BENOEM je plaatje/CATEGORISEER je plaatje*) were displayed on the left and right sides of the screen, and the text *SAME picture* or *DIFFERENT pictures* (*HETZELFDE plaatje/VERSCHILLENDE plaatjes*) was displayed at the top of the screen. In one-third of the blocks, both participants named, while in the other two-thirds, one participant named while the other categorized; this meant that each participant performed the naming task on two-thirds of trials. A condition in which both participants categorized the items was not included because the goal of the experiment was to examine naming latencies and in such a condition, neither of the participants would contribute a naming latency. Task was crossed with object type such that in three blocks, participants saw the same objects, and in three blocks, they saw different objects, which were from different semantic categories to exclude any semantic relatedness effects. Block order was counterbalanced across lists, such that each type of block appeared in each position of the experiment; two randomized orders of items were created for each block to make twelve experimental lists.

2.2.4. Procedure

In the partner-present condition, two participants were simultaneously invited to the lab and introduced to each other. In the partner-absent condition, one participant came to the lab and was told that while the task would normally involve a partner, their partner was not able to attend the experiment and they would therefore be taking part alone. Other than this, the experimental setup, task, and apparatus were identical between conditions, with an experimental script always presenting stimuli to both computers.

The experiment began with a familiarization phase in which participants were shown the 30 objects and their names and categories (printed underneath). Objects were displayed centrally for 4000 ms

each with 200 ms inter-stimulus intervals. Next, participants were asked to put on a unidirectional microphone headset and a pair of noise-cancelling headphones, which played continuous Pink noise in order to mask the partner's response. Participants were then seated against a forehead rest to maintain position about 90 cm from the monitor. After calibration, there were six 'same object' practice trials consisting of three trials in which both participants named *apple*, *bicycle*, and *hat*, and three in which both participants categorized *snake*, *table*, and *carrot*. This practice phase also served to test the noise-cancelling headphones; at the end of the practice phase, the Pink noise volume was adjusted as needed to mask the partner's response.

Following the practice phase were the six blocks of the experiment. Before each block, participants saw a screen which informed them whether the pictures would be the same or different, and which instructed each participant to either name or categorize their picture. Trials began with fixation points on the appropriate sides of the screen for each participant; the experimenter manually registered each participant's fixation. When both participants' fixations were registered, the trial sequence was triggered. It consisted of a blank screen, shown for 500 ms, then the stimuli, shown for 1000 ms, and then another blank screen, shown for 900 ms. Between blocks, a blank screen was presented for five seconds. The experimental session lasted approximately 20 min.

2.2.5. Apparatus

Participants were seated side-by-side in front of identical 24" display monitors (1920 × 1080 resolution). On each monitor, two stimuli were presented on opposite sides of the screen with a center-to-center distance of 36 cm (22.62° visual angle); this distance combined with the visual angle of the stimuli (3.39°) meant that when fixating one picture, participant were highly unlikely to be able to identify the other picture using peripheral vision. Stimulus presentation was controlled by SR Research Experiment Builder software (Version 1.10.1630) on a BenQ computer. Pink noise was played through an Olympus LS-10 Linear PCM recorder connected to two pairs of noise-cancelling headphones (Audio Technica ANC70 Quiet Point) and speech was recorded using a Shure SM10A head-mounted microphone.

Eye movements were tracked with an eye-tracking system including

two host computers, two trackers (EyeLink 1000 Plus Desktop Mount, Version 5.09), and two display computers. Throughout the experiment, the host computers recorded participants' fixations and synchronized the display to the two participants. The right eye of each participant was tracked at 500 hertz with a spatial accuracy of about 0.25° to 0.5°. Areas of interest were defined as the 200 × 200 pixel square containing each object.

2.2.6. Data analysis

Utterance onset was analyzed for all trials in which participants correctly named or categorized their object. Latencies were defined as the time from stimulus onset to response onset; these were hand-coded by the second author using Praat (Boersma & Weenink, 2017). Data were analyzed with mixed-effects models using the *lme4* package, version 1.1–18 (Bates, Maechler, Bolker, & Walker, 2015) in R version 3.5.1 (R Core Team, 2018). Primary dependent measures were log-transformed naming latencies and odds of fixating one or more times on the partner object during naming trials. Log-transformed categorization latencies and odds of fixating one or more times on the partner object during categorizing trials served as secondary dependent measures and are reported in Appendix B. Predictors were partner presence (present/absence), partner task (name/categorize), and picture congruence (same/different), all contrast coded as – 0.5 and 0.5.

All models were initially fit with a maximal random effects structure, including random intercepts by participant, speaker picture and partner picture, and random slopes by participant, speaker picture and partner picture for partner presence, partner task and picture congruence and their interactions. This structure was simplified due to non-convergence to reflect the maximal random effects structure justified by the data by removing the higher-order terms that accounted for the least variance, removing interactions first and then main effects as necessary (see Barr, Levy, Scheepers, & Tily, 2013). Random slopes correlated above 0.9 were also removed to avoid overfitting. For linear models, *p*-values were obtained by model comparison.

2.3. Results

Participants fluently produced the intended response on 97% of the naming trials. The remaining 3% (538 trials, of which 56% were in the partner-present condition) were excluded from further analysis. On these trials, the participant performed the wrong task (90 trials), used a non-target name (187 trials), produced a disfluent response (217 trials), or produced no response (44 trials). No statistical analyses of these low error rates were performed.

2.3.1. Naming latencies

As shown in Fig. 2, naming latencies were affected by partner task in the form of interactions between partner task and the other two variables. In the partner-present condition, trials on which the partner was categorizing elicited a slower average naming latency ($M = 723$ ms) than trials on which the partner was naming ($M = 707$ ms). In the partner-absent condition, this was not the case (partner-categorizing: $M = 713$ ms, vs partner-naming: $M = 722$ ms). In addition, partner-naming trials on which the pictures differed elicited a slower average naming latency ($M = 723$ ms) than trials where the pictures were identical ($M = 706$ ms); for partner-categorizing trials, a trend in the opposite direction was seen (same: $M = 723$ ms, vs different: $M = 713$ ms). These patterns were confirmed statistically in a linear mixed effect model (see Table 1). Critically, no other main effects or interactions were reliable: naming latencies were statistically equivalent in the partner-present ($M = 715$ ms) and partner-absent ($M = 717$ ms) conditions, as well as in the same-picture ($M = 714$ ms) and different-picture ($M = 718$) conditions.

2.3.2. Partner fixations

Most trials had no fixations to the partner object, with one or more

fixations to the partner object on only 7.7% of trials. First, we tabulated whether these fixations to partner pictures occurred before or after onset of speech by merging the eye-tracking fixation data with the onset latencies coded for each trial. This was done because information gained from fixations to partner objects could only affect naming latencies if the fixation occurred prior to speech onset. Individuals fixated the partner object one or more times before the onset of their own speech on only 1.36% of trials. This means that the large majority of the fixations to the partner object occurred after speech onset, indicating that overt visual attention to partner objects could not be the source of interference in naming.

For naming trials, we also examined how odds of fixating the partner object at any point in the trial were modulated by experimental condition. These results are displayed in Fig. 3. Regardless of partner presence, individuals were more likely to fixate the partner object on partner-naming than on partner-categorization trials (9.3%, vs 6.5% of trials). Moreover, regardless of partner presence, individuals fixated the partner object more often when it was different from their own object rather than identical (9.0% vs 7.7% of trials). These patterns were confirmed with a logistic mixed effect model (see Table 2). The fitted model contained only main effects and no interactions; due to the sparse number of fixations to partner objects, a model containing all main effects and interactions failed to converge.

2.4. Discussion

In Experiment 1, we investigated how participants' naming latencies and gaze patterns were affected by the physical co-presence of a task partner, by the match or mismatch of the participants' tasks, and by the match or mismatch of the two partners' stimuli. We did not observe a significant main effect of partner presence on any dependent measure: speakers did not differ in speed or accuracy based upon partner presence, nor did they look more often at the partner's object when the partner was present than when they were absent. This is consistent with earlier studies in our lab indicating that participants carry out simple picture naming tasks in a similar fashion in monologue and dialogue situations (Hoedemaker & Meyer, 2019; Meyer, Alday, Decuyper, & Knudsen, 2018).

We predicted that if individuals co-represent their partner's task, there should be an interaction between partner presence and partner task, and if individuals co-represent their partner's speech content, there should be an interaction between partner presence and picture congruence. We observed the predicted interaction of partner presence and partner task. In the partner-present condition, participants were faster to name pictures when the partner was also naming than when they were categorizing. This indicates that participants represented the partner's actions, leading to interference in naming when the actions were different. While the presence of an effect of partner task replicates Gambi et al. (2015), the direction of the effect was reversed. This discrepancy could have arisen for a variety of reasons, including the relative ease of our task or the relative ease of naming versus categorizing. We elaborate more upon these points in the General discussion section.

We did not observe the predicted interaction of partner presence and picture congruence. There was no main effect of picture congruence either, but there was an interaction between picture congruence and partner task: Naming latencies were only faster in the same-picture than in the different-picture condition on partner-name trials, but not on partner-categorize trials. This follows from the fact that it was only when both task and picture were shared that participants produced the same responses: when the tasks differed, participants produced different responses (e.g. saying "living" and "rabbit" to the same object) regardless of whether their pictures matched. Importantly, this pattern held regardless of partner co-presence. This suggests that even in the partner-absent condition, participants represented to some extent what the partner would have said had they

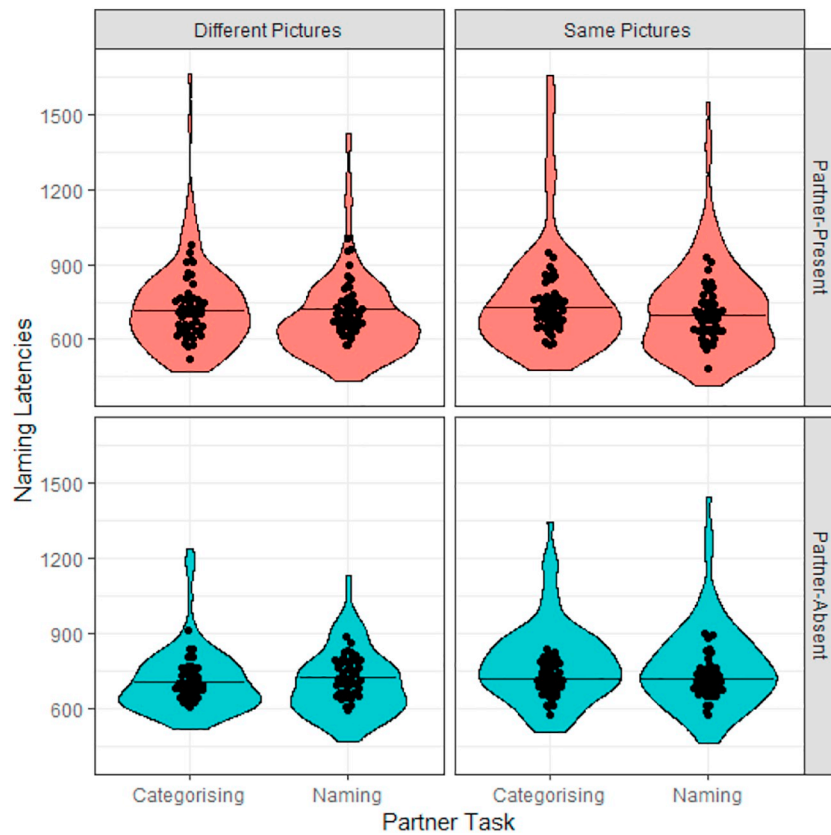


Fig. 2. Distribution of naming latencies by trial in Experiment 1 split by picture congruence (same, different), partner presence (present, absent) and partner task (categorizing, naming). Points represent means by participant and horizontal lines reflect overall condition means.

Table 1
Linear mixed effect model for log-transformed naming latencies in Experiment 1. In the random effects tier, random slopes are nested below the corresponding intercept.

Fixed effects	Estimate	SE	t value	p(χ^2)
Intercept	6.551	0.021	314.5	< 0.001
Partner Presence (Present vs Absent)	-0.007	0.020	-0.35	0.73
Partner Task (Categorizing vs. Naming)	-0.004	0.007	-0.62	0.53
Picture Congruence (Same vs. Different)	0.006	0.006	1.11	0.27
Partner Presence \times Partner Task	-0.029	0.013	-2.17	0.03
Partner Presence \times Picture Congruence	0.011	0.012	0.93	0.36
Partner Task \times Picture Congruence	0.034	0.012	2.76	0.01
Partner Presence \times Partner Task \times Picture Congruence	0.018	0.024	0.73	0.46

Random effects	Term	Variance
Speaker Picture	Intercept	0.0098
Partner Picture	Intercept	0.0001
Participant	Intercept	0.0095
	Partner Task	0.0033
	Picture Congruence	0.0023
	Partner Task \times Picture Congruence	0.0101
Residual		0.0296

been present. The implication is that the expected content of a partner utterance may cause some interference in naming, even when no partner is physically co-present.

Analyses of the participants' eye gaze showed that they largely focused on their own item: fixations to the partner's item were rare and mostly occurred after utterance onset. This provides evidence that the interference effect observed on naming latencies did not rely on the deployment of overt visual attention to the partner picture. Given the

large visual angle of the two stimuli, it would have been very difficult for the participants to identify the partner objects without fixating upon them. Combined with the fact that participants could not hear their partner speak, the near absence of gazes to the partner objects before speech onset means that interference on naming latencies was not based on knowledge of the specific word the partner would say on a given trial. Instead, interference may derive from knowing and representing the partner's task (naming or categorizing the same or different pictures), as this information was provided in the block instructions.

Throughout the whole duration of the trial, partner object fixations were more frequent in the different-picture condition, where the picture provided new information to the participant, than in the same-picture condition, suggesting that overt visual attention to partner objects was deployed more often when it could yield novel information. Fixations to the partner's object were also slightly more frequent on trials where both participants were naming, compared to trials where the partner was categorizing. It is not clear why this difference arose. One might speculate that the participants perceived the partner-naming condition more strongly as a joint task, akin to speaking in chorus, and consequently considered the partner's object to be more task-relevant. It is also possible that the partner-naming task condition involved a simpler representation of the joint task (e.g., she is doing the same thing that I am doing), meaning that participants could more easily allocate spare attentional resources to their partner's object. Both possibilities are consistent with the conceptualization of language as a form of joint action.

In sum, Experiment 1 showed evidence for interference in naming latencies based upon the partner's task in the partner-present condition, supporting the actor co-representation account of Gambi et al. (2015) and suggesting that the co-presence of a partner leads to interference based upon expectations about their actions. Experiment 1 also showed reduced interference when the two partners were producing identical

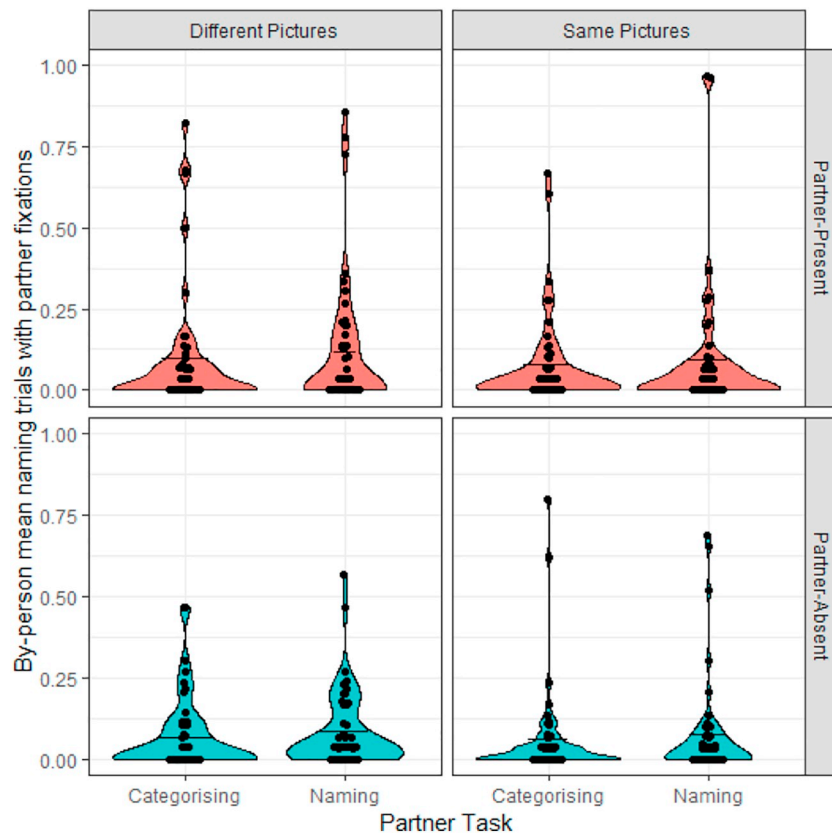


Fig. 3. Distribution of average proportions of fixations to partner pictures by participant in Experiment 1, split by partner presence (present, absent), picture congruence (different, same) and partner task (categorizing, naming). Points represent means by participant; horizontal lines reflect condition means.

Table 2
Logistic mixed effect model for odds of fixating the partner object in Experiment 1 naming trials.

Fixed effects	Estimate	SE	z value	p(z)
Intercept	-3.34	0.19	-17.90	< 0.001
Partner Presence (Present vs Absent)	0.25	0.35	0.71	0.48
Partner Task (Categorizing vs. Naming)	0.29	0.08	3.81	< 0.001
Picture Congruence (Same vs. Different)	0.22	0.08	2.84	0.01

Random effects	Term	Variance
Speaker Picture	Intercept	0.00
Partner Picture	Intercept	0.08
Participant	Intercept	2.60

responses (the partner-name, same-picture condition), indicating that some aspects of the content of the partner's speech were also mentally represented, though this was not contingent on partner co-presence.

3. Experiment 2

3.1. Introduction

In Experiment 1, we observed that partner task affected naming latencies, such that participants were faster to name pictures on partner-name than partner-categorize trials. In this experiment, the categorization responses were verbal responses that classified the object as living or non-living. Therefore, it is not clear whether the task-congruence effect arose because participants represented the partner's task at an abstract conceptual level (naming or categorizing) or because they broadly represented the linguistic content of their utterances (an adjective referring to a property of the object or noun referring to a basic

level term). To distinguish between these options, Experiment 2 assessed whether similar patterns of interference occurred when the verbal categorization task used during Experiment 1 was replaced by an equivalent forced-choice button-press task.

Previous work has demonstrated robust social interference effects based upon the expected button-press of a partner (the 'Social Simon Effect'; e.g., Sebanz, Knoblich, & Prinz, 2003), making a button-pressing categorization task ideal for examining socially situated interference. If interference arose from a predicted verbal response that was incompatible with the participant's own verbal response, interference should be minimal in partner-categorize trials when the partner's response is non-verbal. In contrast, if what matters is purely the conceptual properties of the partner's response (name versus categorize), robust interference in naming should still be observed in partner-categorize trials even when the partner is producing a non-verbal response.

To simplify the experimental design, Experiment 2 included only the partner-present condition. As in Experiment 1, the two participants again either saw the same or different pictures. This allowed us to evaluate whether naming latencies, accuracy, or eye gaze were affected by picture congruence, providing a partial replication of Experiment 1. The hypothesis was that if individuals represent their partner's task even if it is non-verbal, we would observe a main effect of partner task. If individuals represent their partner's response even if it is non-verbal, we predict a main effect of picture congruence, and if individuals represent their partner's response only if it is verbal, we predict an interaction between partner task and picture congruence.

3.2. Method

3.2.1. Participants

Based upon the power analyses reported in Experiment 1, a sample size of 48 participants was selected. This reflects the fact that we

removed the partner-absent condition; as such, this sample would still allow us to an effect f of 0.1 or larger reliably. 58 participants between the ages of 18 and 30 years old (mean age 23) were recruited from the Max Planck Institute for Psycholinguistics' database. Ten participants were excluded due to calibration issues, track loss, or poor-quality audio recordings. Categorization data for two participants were lost due to a keyboard malfunction; the naming data from these individuals were included in the reported analyses.

The final sample of 48 participants (37 female) reported normal or corrected-to-normal vision and no speech or hearing problems, and received a payment of 8 € for their participation. Informed consent was obtained from each participant. The study was approved by the ethics board of the Faculty of Social Sciences of Radboud University.

3.2.2. Materials and design

The same materials and design were used as in Experiment 1, except that all participants were tested in the partner-present condition. In two pairs of participants there was one individual who could not be calibrated. These two participants were excluded (as noted above) and the data from the remaining individual in both pairs was left in the experiment. These exclusions led to a small mismatch in counterbalancing across experimental lists: the final data set contained data from five individuals in one list, three in another, and four each in the remaining ten lists.

3.2.3. Apparatus and procedure

The apparatus and procedure were the same as in Experiment 1, except that participants made categorization judgments by pressing the F key (for living) or J key (for non-living) on a computer keyboard. Response latencies for this measure were collected during the 1000 ms interval for which the images remained present on the screen. Participants also completed a 30 question Big IV inventory (Soto & John, 2017; unpublished Dutch version translated, validated, and shared by Jaap Denissen, personal communication) used in exploratory analyses that are not reported here.

3.2.4. Data analysis

As in Experiment 1, with the omission of the partner presence predictor. Transcriptions of sound files and annotations of onset latencies for each utterance were performed by trained research assistants and checked by the first author.

3.3. Results

Participants fluently produced the intended response on 97% of naming trials. The remaining 3% (194 trials) were excluded from further analysis. On these trials, the participant produced no response (58 trials), used a non-target name (126 trials), or produced a disfluent or incomplete response (10 trials).

3.3.1. Naming latencies

We observed a main effect of partner task such that naming times were slower on trials on which the partner was categorizing ($M = 722$ ms) than naming ($M = 702$ ms), see Fig. 4. This was supported by a linear mixed effect model (see Table 3). No other effect was reliable: Picture congruence did not have a significant main effect (Different pictures $M = 718$ ms; Same pictures $M = 706$ ms), and the interaction between partner task and picture congruence was also not significant, with minimal differences in latencies between conditions (Partner-name, same-picture $M = 697$ ms; Partner-name, different-picture $M = 708$ ms; Partner-categorize, same-picture $M = 715$ ms; Partner-categorize, different-picture $M = 729$ ms).

3.3.2. Partner fixations

Participants fixated on the partner object on only 7.5% of trials. As described in Experiment 1, we again tabulated whether looks to the

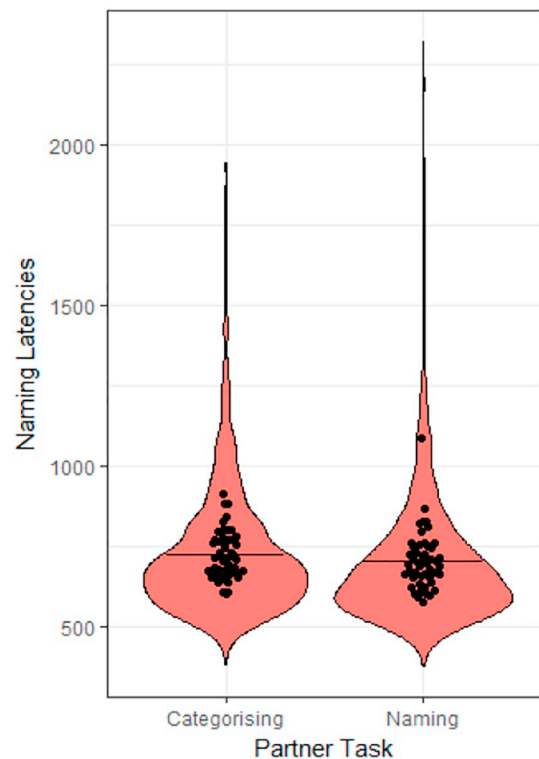


Fig. 4. Distribution of naming latencies by trial in Experiment 2 split by partner task. Points represent means by participant; horizontal lines reflect condition means.

Table 3

Linear mixed effect model for log-transformed naming latencies in Experiment 2. In the random effects tier, random slopes are nested below the corresponding intercept.

Fixed effects	Estimate	SE	t value	$p(\chi^2)$
Intercept	6.542	0.025	257.13	< 0.001
Partner Task (Categorizing vs. Naming)	-0.032	0.012	-2.66	0.01
Picture Congruence (Same vs. Different)	0.016	0.010	1.58	0.12
Partner Task \times Picture Congruence	-0.011	0.010	-1.05	0.29

Random effects	Term	Variance
Speaker Picture	Intercept	0.013
	Partner Picture	0.000
	Picture Congruence	0.001
Participant	Intercept	0.009
	Partner Task	0.006
	Picture Congruence	0.003
Residual		0.035

partner picture occurred before or after onset of speech. As in Experiment 1, individuals rarely fixated the partner object before the onset of their own speech, with fixations to partner objects before speech onset occurring on only 0.54% of trials. This provides evidence that the interference observed in naming latencies did not derive from the allocation of overt visual attention to partner objects.

We also again examined how the odds of fixating partner objects at any point in the trial were modulated by condition. As shown in Fig. 5, individuals were significantly more likely to fixate the partner object during naming trials when the partner was also naming (10.3%, vs 6.6% for partner-categorize trials) and when the partner object differed from their own (9.4%, vs 7.5% for same-item trials), with no reliable interaction between the two factors. These patterns were confirmed with a logistic mixed-effect model (see Table 4).

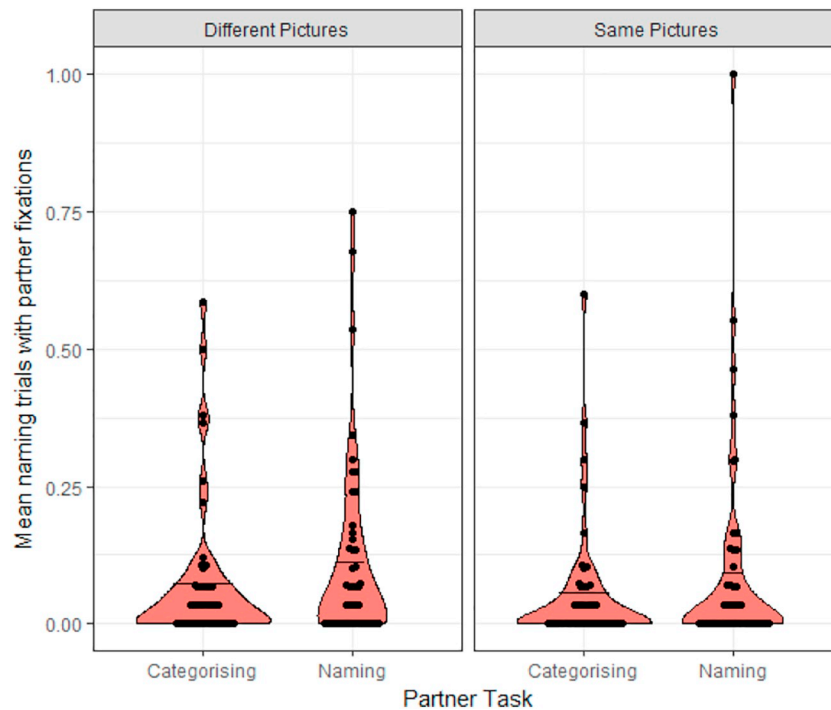


Fig. 5. Distribution of average proportions of fixations to partner pictures by participant in Experiment 2, split by picture congruence (different, same) and partner task (categorizing, naming). Points represent means by participant; horizontal lines reflect condition means.

Table 4
Logistic mixed effect model for odds of fixating the partner object in Experiment 2 naming trials.

Fixed effects	Estimate	SE	z value	p(z)
<i>Intercept</i>	-3.47	0.29	-12.02	< 0.001
<i>Partner Task (Categorizing vs. Naming)</i>	0.49	0.11	4.52	< 0.001
<i>Picture Congruence (Same vs. Different)</i>	0.24	0.11	2.23	0.03
<i>Partner Task x Picture Congruence</i>	0.07	0.22	0.31	0.76

Random effects	Term	Variance
<i>Speaker Picture</i>	<i>Intercept</i>	0.01
<i>Partner Picture</i>	<i>Intercept</i>	0.06
<i>Participant</i>	<i>Intercept</i>	3.21

3.4. Discussion

In Experiment 2, we investigated how participants' naming latencies and overt visual attention to simple line drawings were affected by the match or mismatch of the participants' tasks (both participants naming, or one participant naming while the partner categorized pictures using a button-press task) as well as by the match or mismatch of the two partners' pictures. We predicted that if individuals represent their partner's task even if it is non-verbal, we would observe a main effect of partner task. In addition, we predicted that if individuals represent their partner's response regardless of whether it is verbal, there would be a main effect of picture congruence, and if individuals represent their partner's response only if it is verbal, there would be an interaction between partner task and picture congruence.

Replicating Experiment 1, we observed the expected effect of task congruence on naming latencies, which were faster on trials where the partner was also naming than on trials where the partner was performing the categorization task. This again provides support for the view that individuals mentally represent their partners' task regardless of whether it requires a verbal response, and that this leads to interference in naming when the partner's task is different.

We did not observe the predicted effect of picture congruence on naming latencies. Naming times were numerically faster in the same-picture than in the different-picture condition, but this 12 msec difference did not reach significance. In Experiment 1, an effect of picture congruence had been observed, with latencies being shorter in the same-picture same-task condition (the only condition where both speakers produced the same word) than in the remaining conditions. Together, these results suggest interference may sometimes arise due to the content of a partner's speech, though the effect appears to be less robust than the effect of partner task.

In Experiment 2, we again observed few fixations to partner objects, and most of these occurred after the onset of speaking. Replicating Experiment 1, we observed that the likelihood of fixating the partner object was higher when the partner picture was different from the participant's own picture. We also again observed that the likelihood of fixating the partner object was higher when both participants were performing the same task than when they were engaged in different tasks. This is again consistent with the characterization of the joint naming task as a form of joint action, such as naming in chorus.

4. General discussion

In two experiments, participants named simple line drawings while a partner seated next to them was engaged in either the same or a different task pertaining to either the same or a different picture. This was a simple, fundamentally non-communicative paradigm that could easily be performed without any knowledge of the partner's task or picture, yet we observed robust interference in naming latencies from the partner's task in both experiments. Participants consistently named objects faster when their partner was also naming objects than when their partner was categorizing them, no matter whether the categorization was performed verbally (Experiment 1) or non-verbally (Experiment 2). In Experiment 1, we also observed significant interference in naming based upon the partner's speech content. Task congruence effects like those observed in both experiments have been observed in other contexts and are widely discussed in theories of joint action (e.g. Knoblich et al., 2011; Sebanz et al., 2003, 2006; Vesper

et al., 2017). The implication is that our pairs of participants conceptualized their task as a joint one by drawing on their everyday experience with conversation, despite being fundamentally able to carry out the task entirely independently of their partner. This supports the premise suggested in earlier work (e.g. Clark, 1996; Gambi et al., 2015; Pickering & Garrod, 2004, 2013) that language is a form of joint action. As such, the mechanisms that allow us to co-represent partner speech affect language, but may not be language-specific.

The results of the current study are consistent with those obtained by Gambi et al. (2015), who also found strong evidence for representation of the partner's task with weaker evidence for representation of specific utterances. This dissociation between representation of the task and of the content of the partner's speech was what motivated us to run a similar study. Our expectation had been that, in a simpler paradigm where the partners were physically present in the same room, more evidence for the co-representation of speech content might be found. This expectation was not confirmed: we again found limited evidence for the co-representation of partner speech content. In combination with other results from joint language paradigms (e.g. Corps, Crossley, Gambi, & Pickering, 2018; Hoedemaker et al., 2017; Hoedemaker & Meyer, 2019; Kuhlen & Rahman, 2017), this suggests that while individuals can represent what the partner will say in fine-grained detail, partial co-representation may often be a useful default.

Our results point to an important contrast between language and some of the other joint tasks examined in the literature. For many joint tasks, one might expect participants to represent their partner's task and behavior in as much detail as possible. For example, robust interference is often observed when partners respond to different stimuli or produce different responses, leading to joint versions of effects like the Simon effect (e.g. Sebanz et al., 2003) and the flanker effect (Atmaca, Sebanz, & Knoblich, 2011). The implication is that while individuals are capable of representing their partner's behavior in joint tasks in much detail, the language tasks examined so far have elicited only partial co-representation. We speculate that the participants' tendency to resort to partial co-representation may come from their experience with conversation. Awareness of the interlocutor and the fact that they are likely to speak may be foundational for smooth turn-taking, but generating precise predictions about the content of a partner's utterances may be dysfunctional, as it would interfere with a speaker's own speech planning processes (see Hoedemaker & Meyer, 2019, for similar arguments.) Other forms of joint action may also be best supported by partial co-representations. Consider the joint tasks of lifting a piano versus playing a piano duet. Precise predictions about the partner's movements to the millimeter and timing to the millisecond are unnecessary to lift a piano in synchrony, but might be required for playing it in synchrony. We leave the question of which other forms of joint action lead to partial co-representations for future work.

4.1. Comparison to other joint naming studies

Though our findings are consistent with the main results reported by Gambi et al. (2015), two differences merit brief discussion. While we observed an effect of task interference, as in Gambi et al. (2015), the direction of this effect was reversed. In their Experiment 2, Gambi et al. (2015) observed more interference for partner-name than partner-categorize trials. (The remaining experiments in their study did not include a categorization task, but instead featured a partner no-response condition). We consistently observed the opposite, with increased interference for partner-categorize trials observed in both Experiments 1 and 2. The reason for this reversal might be methodological. First, note that the studies differed both in the naming and categorization tasks: The participants in Gambi et al. (2015) named two objects or categorized whether or not they belonged to the same semantic category. By contrast, the participants in our experiments named only one object or categorized it as being animate or inanimate. Due to these task

differences, it is possible that there was a relative benefit from representing a shared task in our study, or less interference derived from representing different tasks. Further experiments including an additional partner no-task condition would be necessary to distinguish these possibilities. We also note that our task was less demanding than the task in Gambi et al. (2015), which could have led to an overall reduction of interference based on the partner object's name; this could also potentially have contributed to the reversal of the effect. In sum, though the reversal of the task effects between the two studies was unexpected, it is important to note that both convincingly showed that the participants represented their partners' tasks.

A second difference between our studies and those of Gambi et al. (2015) is that we found more evidence than they did for the representation of the partner's speech content. This resulted in a significant interaction in Experiment 1, where we found that when both participants were producing the same response (the same-picture, partner-name condition), naming latencies tended to be the fastest. In Experiment 2, the experimental conditions patterned numerically in the same direction, with the fastest naming latencies in the same-picture, partner-name condition but the effect was not significant. This means that despite some superficial differences, the overall pattern of content interference still matches Gambi et al. (2015), who found that the content of partner speech occasionally influenced error rates.

The consistent effects of task co-representation and weaker effects of content co-representation observed in our experiments and in Gambi et al. (2015) are informative when placed against work by Corps et al. (2018). This study used a communicative joint naming task in which individuals had to respond to a partner utterance by either answering a question ("Have you visited the city of Paris?") with a yes/no response or by using a key-press to predict when the end of the question was upcoming. When participants had to respond to the utterance content by answering the questions, the latencies were affected by the predictability of the question-final words, demonstrating the co-representation of utterance content. However, when participants only had to respond to end of the utterance, content predictability did not matter, demonstrating that a full co-representation of partner content is not always formed. This pattern highlights the fact that representations of content can be recruited and used when needed. In other words, the degree of co-representation is likely to be flexible such that individuals can and do predict partner content when it is important for their own speech plan.

4.2. Representations of partner utterances and overt visual attention

We collected eye-tracking data in order to examine the link between interference in naming and visual attention. In both experiments, we found more interference in naming latencies when the partner was performing a different task, while attention to partner objects followed the opposite pattern. Given the visual angle between the two pictures, it is unlikely that participants could have attended to the partner image without directly fixating it. This means that interference in naming latencies did not rely on overt visual attention to the partner object.

Instead, we found that participants looked infrequently at the partner object, and did so most often after producing their own response. This means that fixations to partner objects did not reflect visual attention recruited for speech planning (contra most usages of eye-tracking in language production research). We speculate that since looks to the partner object occurred more often when the joint task was easy (in the partner name condition) and when there was new information to be gained (in the different picture condition), fixations to partner objects might have occurred simply because time remained at the end of the trial and there were only two possible pictures to fixate. One might also speculate that fixations to partner object may have been more frequent in the same task than in the different task condition because in the former condition participants felt more strongly that they were engaged in a joint task with the partner.

4.3. Joint naming and conversation

To situate our findings within the larger literature, we note that arguably the largest contribution to the study of dialogue follows from a conversation analysis perspective (e.g. Levinson & Torreira, 2015; Sacks, Schegloff, & Jefferson, 1978). This work has demonstrated that individuals have the ability to tightly coordinate their utterances in naturalistic conversations, implying that in order to converse fluently, individuals must co-represent some aspects of their partner's utterances. The present work, as well as earlier related work (e.g. Corps et al., 2018; Hoedemaker et al., 2017; Hoedemaker & Meyer, 2019; Kuhlen & Rahman, 2017; Meyer et al., 2018) departed from a naturalistic paradigm to use a classic experimental psycholinguistic design (simple, normed items presented in randomized blocks that are in turn presented in a counterbalanced order)—in essence, treating language as a simple form of action. This strand of work illustrates the role for psycholinguistic laboratory studies for understanding conversation, and we hope that our work inspires future studies in this domain. We believe that by combining experimental studies with more naturalistic work from a conversation analysis perspective, we can gain a clearer understanding about what individuals and dyads do when they talk to each other.

We also suggest that this work further demonstrates the need to consider dyadic speech in psycholinguistic processing theories. In conversations, individuals do not produce parallel monologues, but

represent aspects of their interlocutors' plans. While most research on language production focuses on single individuals producing de-contextualized utterances, understanding how individuals balance the needs to listen and plan simultaneously provides insight into basic mechanisms of language production. The implication from our work and from other joint naming tasks is that partner utterances can lead to interference for multiple reasons; this means that speaking in dialogue may differ both in attentional control processes and in speech planning processes from speaking in monologue.

5. Conclusion

Replicating earlier findings by Gambi et al. (2015), we found that dyads of speakers strongly represent whether both individuals are engaged in the same or different tasks, leading to interference in naming latencies. However, individuals did not seem to strongly represent whether they both saw the same or different stimuli or responded in the same or different ways. This provides evidence that co-representations of partner speech may be coarse-grained and partial. We propose that the use of such coarse partner representations may arise from the functionality of coarse representations of partner's speech in everyday conversation. To take turns successfully, one needs to expect that the conversation partner will also want to speak, but to listen successfully, one does not need to form detailed expectations of what is going to be said.

Appendix A. Items

Category	Dutch name	English name	Visual complexity	Log frequency	Number syllables	Name agreement	Normed RT
Non-living	Broek	Pants	16.6	1	1	91	804
Non-living	Hoed	Hat	9.87	1.61	1	100	862
Non-living	Trui	Sweater	12.4	1.30	1	77	977
Non-living	Jas	Coat	14.5	1.69	1	82	1067
Non-living	Rok	Skirt	10	1.49	1	85	1065
Non-living	Tafel	Table	12.9	2.39	2	100	734
Non-living	Bed	Bed	14.5	2.48	1	97	830
Non-living	Stoel	Armchair	12.4	1.18	2	73	1017
Non-living	Kast	Dresser	21.3	1.68	1	81	1082
Non-living	Bureau	Desk	18.2	1.94	2	87	1118
Non-living	Auto	Car	10.4	2.32	2	100	694
Non-living	Fiets	Bike	24.4	1.68	1	100	759
Non-living	Vliegtuig	Plane	17.3	1.72	2	100	831
Non-living	Bus	Bus	23.3	1.76	1	100	955
Non-living	Boot	Boat	17.7	1.83	1	87	1061
Living	Slang	Snake	23.9	1.43	1	100	772
Living	Olifant	Elephant	24.5	1	3	100	788
Living	Konijn	Rabbit	12.2	1.36	2	100	813
Living	Muis	Mouse	14	1.32	1	86	997
Living	Bij	Bee	13	3.58	1	54	1252
Living	Appel	Apple	9.37	1.38	2	100	856
Living	Citroen	Lemon	9.68	1.04	2	97	901
Living	Peer	Pear	19.3	1	1	100	914
Living	Aardbei	Strawberry	17.3	0.70	2	100	917
Living	Kers	Cherry	5.8	0.70	1	89	1019
Living	Paddestoel	Mushroom	9.51	0.95	3	95	872
Living	Wortel	Carrot	14.1	1.57	2	94	885
Living	Tomaat	Tomato	9.57	0.95	2	97	912
Living	Sla	Lettuce	17.6	0.48	1	66	986
Living	Mais	Corn	16.7	0.90	1	92	1005

Appendix B. Categorization latencies and fixation patterns

Both experiments included conditions where both partners named the pictures and conditions where one of them named the picture, while the partner categorized it. The design omitted the condition where both participants categorized the pictures. Therefore, we cannot assess the task interference effect on performance in the categorization task. Other results from the categorization conditions are reported below.

Experiment 1

There were no reliable differences in mean categorization latencies based on any predictors in a linear mixed effect model (Table B1), though

there was a numerical trend for slower categorization latencies in the partner-present condition (752 ms) than the partner-absent condition (720 ms). Individuals were significantly more likely to fixate the partner object when it differed from their own (7.5%, versus 5.4% for same-picture trials). This was confirmed with a logistic mixed effect model (see Table B2); no other main effects or interactions were reliable.

Experiment 2

Participants only provided a response in the recorded time window for 76% of categorization trials, with the correct response given on 1975 trials (72%) and the incorrect response given on 121 trials (4%). On 658 trials (24%), no latency was recorded, indicating a response time longer than 1000 ms or no response, and on 6 trials, the participant did the wrong task. We attribute the large amount of missing data to a coding error that recorded latencies only for the 1000 ms that the picture was present on the screen.

For the recorded response latencies, there were no reliable differences between conditions in a linear mixed effect model (see Table B3). Individuals were numerically more likely to fixate the partner object when it differed from their own (8%, versus 3% for same-picture trials) but this was not significant in a logistic mixed effect model, nor were any other main effects or interactions (see Table B4).

Table B1

Linear mixed effect model of log-transformed categorization latencies in Experiment 1. For these trials, the partner was always naming.

Fixed effects	Estimate	SE	t value	p(χ^2)
<i>Intercept</i>	6.579	0.014	474.90	< 0.001
<i>Partner Presence (Present vs Absent)</i>	0.035	0.026	1.34	0.18
<i>Picture Congruence (Same vs. Different)</i>	-0.008	0.004	-1.70	0.09
<i>Partner Presence × Picture Congruence</i>	0.001	0.009	0.17	0.87
Random effects	Term			Variance
<i>Speaker Picture</i>	<i>Intercept</i>			0.000
<i>Partner Picture</i>	<i>Intercept</i>			0.000
<i>Participant</i>	<i>Intercept</i>			0.016
<i>Residual</i>				0.028

Table B2

Logistic mixed effect model for odds of fixating the partner object in Experiment 1 categorizing trials. For these trials, the partner was always naming.

Fixed effects	Estimate	SE	z value	p(z)
<i>Intercept</i>	-4.18	0.28	-14.83	< 0.001
<i>Partner Presence (Present vs Absent)</i>	0.11	0.48	0.24	0.81
<i>Picture Congruence (Same vs. Different)</i>	0.42	0.13	3.19	< 0.01
<i>Partner Presence × Picture Congruence</i>	0.24	0.26	0.93	0.35
Random effects	Term			Variance
<i>Speaker Picture</i>	<i>Intercept</i>			0.00
<i>Partner Picture</i>	<i>Intercept</i>			0.29
<i>Participant</i>	<i>Intercept</i>			4.09

Table B3

Linear mixed effect model of log-transformed categorization latencies in Experiment 2. For these trials, the partner was always naming. In the random effects tier, random slopes are nested below the corresponding intercept.

Fixed effects	Estimate	SE	t value	p(χ^2)
<i>Intercept</i>	6.42	0.018	364.6	< 0.001
<i>Picture Congruence (Same vs. Different)</i>	-0.001	0.015	-0.1	0.93
Random effects	Term			Variance
<i>Speaker Picture</i>	<i>Intercept</i>			0.000
	<i>Picture Congruence</i>			0.000
<i>Partner Picture</i>	<i>Intercept</i>			0.001
	<i>Intercept</i>			0.012
<i>Participant</i>	<i>Intercept</i>			0.007
	<i>Picture Congruence</i>			0.027
<i>Residual</i>				0.027

Table B4

Logistic mixed effect model for odds of fixating the partner object in Experiment 2 categorizing trials. For these trials, the partner was always naming. In the random effects tier, random slopes are nested below the corresponding intercept.

Fixed effects	Estimate	SE	t value	p(z)
<i>Intercept</i>	-3.69	0.28	-12.95	< 0.001
<i>Picture Congruence (Same vs. Different)</i>	0.32	0.36	0.89	0.37
Random effects	Term	Variance		
<i>Speaker Picture</i>	<i>Intercept</i>	0.002		
<i>Partner Picture</i>	<i>Intercept</i>	0.000		
<i>Participant</i>	<i>Intercept</i>	2.10		
	<i>Picture Congruence</i>	0.72		

References

- 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.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer [computer software]. Version 6.0.33*.
- Clark, H. H. (1996). *Using language*. 1996. Vol. 952, Cambridge: Cambridge University Press 274–296.
- Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, 175, 77–95.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(1), 1–21.
- 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.
- Hoedemaker, R. S., & Meyer, A. S. (2019). Planning and coordination of utterances in a joint naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(4), 732–752.
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of learning and motivation*. Vol. 54. *Psychology of learning and motivation* (pp. 59–101). Academic Press.
- Kuhlen, A. K., & Rahman, R. A. (2017). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition*, 166, 94–106.
- Levinson, S. C. (2016). Turn-taking in human communication, origins, and implications for language processing. *Trends in Cognitive Sciences*, 20(1), 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>.
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6, 1.
- Meyer, A. S., Alday, P. M., Decuyper, C., & Knudsen, B. (2018). Working together: Contributions of corpus analyses and experimental psycholinguistics to understanding conversation. *Frontiers in Psychology*, 9, 525.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.R-project.org/>.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. *Studies in the organization of conversational interaction* (pp. 7–55).
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–76.
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, 88(3), B11–B21.
- Severens, E., van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, 119(2), 159–187.
- Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, 113(1), 117–143.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634.
- 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.