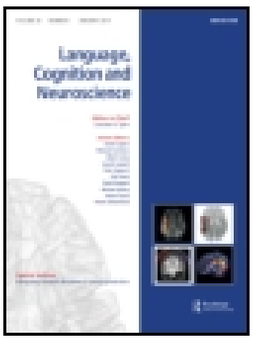

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Next speakers plan word forms in overlap with the incoming turn: evidence from gaze-contingent switch task performance

Mathias Barthel & Stephen C. Levinson

To cite this article: Mathias Barthel & Stephen C. Levinson (2020): Next speakers plan word forms in overlap with the incoming turn: evidence from gaze-contingent switch task performance, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2020.1716030

To link to this article: https://doi.org/10.1080/23273798.2020.1716030

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2020.1716030

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

Next speakers plan word forms in overlap with the incoming turn: evidence from gaze-contingent switch task performance

Mathias Barthel [a,b] and Stephen C. Levinson[a]

[a]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; [b]Humboldt University, Berlin, Germany

ABSTRACT

To ensure short gaps between turns in conversation, next speakers regularly start planning their utterance in overlap with the incoming turn. Three experiments investigate which stages of utterance planning are executed in overlap. E1 establishes effects of associative and phonological relatedness of pictures and words in a switch-task from picture naming to lexical decision. E2 focuses on effects of phonological relatedness and investigates potential shifts in the time-course of production planning during background speech. E3 required participants to verbally answer questions as a base task. In critical trials, however, participants switched to visual lexical decision just after they began planning their answer. The task-switch was time-locked to participants' gaze for response planning. Results show that word form encoding is done as early as possible and not postponed until the end of the incoming turn. Hence, planning a response during the incoming turn is executed at least until word form activation.

ARTICLE HISTORY
Received 3 June 2019
Accepted 27 December 2019

KEYWORDS
turn-taking; speech planning; lexical decision; picture naming; visual world paradigm

1. Introduction

In conversation, interlocutors readily exchange turns of talk, frequently switching from the role of the listener to the role of the speaker without leaving long gaps between turns (Sacks, Schegloff, & Jefferson, 1974; Stivers et al., 2009). Previous studies consistently find that speech planning takes more time than the average gap between turns in conversation, as it takes speakers at least 600 ms to plan single words (Indefrey, 2011) and about one and a half seconds to prepare a simple sentence (Griffin & Bock, 2000; Schnur, Costa, & Caramazza, 2006). Based on evidence from picture naming studies using the picture-word interference paradigm (e.g. Schriefers, Meyer, & Levelt, 1990; Wilshire, Singh, & Tattersall, 2016), time requirements of the separate levels of the speech production process are estimated to be around 200 ms to activate a mental concept that fits a depicted picture, about 75 ms for the selection of a lemma that matches the concept and represents semantic and syntactic information of a word, and approximately 80 ms to retrieve the phonological code of that word (Indefrey & Levelt, 2004), followed by processes of syllabification and phonetic encoding. Recent models of turn taking postulate that next speakers need to start planning their utterance as early as possible (early-planning hypothesis) and in overlap with the incoming turn (Levinson & Torreira, 2015; Pickering & Garrod, 2013), assuming that the gap between turns would be much longer than regularly observed if next speakers only began to plan their turn in reaction to the end of the incoming turn or even to turn-final cues about the upcoming turn end (Barthel, Meyer, & Levinson, 2017). Planning the content of a response turn that is contingent upon the incoming turn can only begin when the incoming message is sufficiently clear or can be reliably anticipated. If response planning is executed in overlap with the incoming turn, the respective planning processes might be slowed down due to concurrent speech comprehension. The time pressures of conversation, the most frequently used speech exchange system, might therefore have a great impact on the mechanisms of speech planning.

Experimental studies testing the early-planning hypothesis have indeed shown that planning commonly begins as early as possible during the incoming turn (but see the study by Sjerps and Meyer (2015), and Barthel, Sauppe, Levinson, and Meyer (2016) for discussion thereof). Barthel et al. (2016) used a list completion paradigm with a confederate listing a number of displayed objects and the participant listing the remaining displayed objects. The confederate turns had different

CONTACT Mathias Barthel mathias.barthel@hu-berlin.de
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2020.1716030.

syntactic structures, so that they either ended with one of the object names or with a verb form that was redundant for participants to plan their next turn. Eye-movements and voice onset latencies showed that participants started to plan their response turn as early as possible during the incoming turn, even if redundant material predictably followed before the incoming turn's end. Bögels, Magyari, and Levinson (2015) used a confederate who asked participants questions whose answer became clear either in the middle of the question or only at the end of the question (e.g. as in "Which character, also called 007, appears in the famous movies?" (early) vs. "Which character from the famous movies is also called 007?" (late)). Response latencies were shorter when the answer could be deduced in the middle of the question than when it became obvious only at its very end. Additionally, in both early and late questions, 500 ms after the onset of the critical information, the authors recorded a positivity in participants' EEG signal, which was substantially reduced in a control task that did not involve response planning. This positivity was therefore interpreted as an indication of early response planning processes. Consistent findings are reported by Corps, Crossley, Gambi, and Pickering (2018). Manipulating the predictability of an incoming question's end, the authors find that participants answered questions earlier when their end was predictable as compared to unpredictable, suggesting that participants used content prediction to begin to plan their answer in overlap with the incoming question whenever possible.

While these studies show that planning starts in overlap with the incoming turn, they did not investigate which levels of production planning are run through while still listening to incoming speech. Using a post-hoc EEG source localisation analysis on the data recorded during their question-answer study, Bögels, Magyari, et al. (2015) found activation of the middle frontal and precentral gyri in overlap with the incoming turn and hypothesised this activation to be due to phonological planning in preparation of the answer. However, these brain regions have also been found to be active during memory retrieval (Rajah, Languay, & Grady, 2011; Raz et al., 2005), which could be responsible for the reported findings instead, since participants needed to retrieve the answers to the posed questions from long term storage. Alternatively, activation of these brain regions might have been due to ongoing comprehension of the incoming question, which is supposed to result in concurrent activation of related speech production processes (Galantucci, Fowler, & Turvey, 2006; Liberman & Mattingly, 1985; Pickering & Garrod, 2013). Hence, the question which stages of response planning are run through in overlap with the incoming turn remains unsettled. Speakers need to go through a number of these stages before being prepared to articulate their turn, including at least conceptualisation, formation of a syntactic structure, lemma selection, word form retrieval, and phonetic encoding (Indefrey & Levelt, 2004; Levelt, 1989). The turn-taking model by Levinson and Torreira (2015) assumes that all stages of response formulation are run through as early as the action that is intended with the incoming turn can be recognised. The model therefore assumes that all the stages of response formulation regularly occur in overlap with the incoming turn, while articulation is withheld until the incoming turn comes to an end. Whether this is true for all stages of speech planning is an open empirical question.

A major reason to assume that some processing stages might be postponed until the end of the incoming turn is the well established fact that speech production and comprehension compete for processing capacities. Previous studies found that incoming linguistic material interferes with speech production more than non-linguistic material, with interference being most severe on the word form level. Kemper, Herman, and Lian (2003) asked participants open questions to elicit free talk while participants continuously performed different secondary tasks. They found that speech production was more difficult for participants when they had to ignore incoming speech than when they had to ignore noise, as was indicated for example by a higher rate of production errors in the speech condition. Schriefers et al. (1990), using the picture word interference paradigm, compared the effect of auditorily presented distractor words with a noise condition and a condition without distractors (silence) on picture naming performance and found that distracting speech was significantly more detrimental to response latencies than silence or noise. Fargier and Laganaro (2016) tested participants on a dual-task with picture naming as base task 1 and either tone or syllable detection as a go/no-go task 2. Analysing only no-go trials, they found that naming latencies were longer with syllables than with tones as concurrent input. Additionally, they found ERP waveform differences between syllables and tones as concurrent input about 400 ms after picture onset, which they interpreted to be caused by increased interference of verbal as compared to non-verbal material with word form encoding processes. Similarly, Fairs, Bögels, and Meyer (2018) found interference on picture naming performance to be larger with a second linguistic tasks (syllable detection) than with a concurrent non-linguistic task (tone identification).[1] Klaus, Mädebach, Oppermann, and Jescheniak (2017) used a dual-task

paradigm asking participants to ignore auditory distractor words and produce subject-verb-object picture descriptions while concurrently performing either a visuospatial or a verbal working memory task. Under verbal but not under visuospatial working memory load, participants' phonological planning scope was reduced to the subject of the sentence, while their abstract lexical planning scope remained unreduced, including the sentence final object. This pattern of results shows that high verbal working memory load interferes with phonological production planning. Taking together these findings, postponing (at least) phonological planning until the end of the incoming turn could therefore be an efficient strategy that might be applied by next speakers to keep the increase in processing costs that come with planning in overlap at a moderate level (Barthel & Sauppe, 2019). On the other hand, late phonological planning might lead to long gaps between turns that might be undesired because long delays give rise to inferences on the turn's meaning (Bögels, Kendrick, & Levinson, 2015; Clayman, 2002; Pomeranz & Heritage, 2012) and might commonly be avoided for that reason.

The present study investigates which stages of formulation (lemma selection and word form retrieval) are executed in overlap with incoming speech, mimicking a situation where a participant of a conversation starts to prepare their own turn while still listening to another person speaking. While planning the next turn in overlap with the incoming turn, each level of processing, from conceptual planning to word form retrieval can be hypothesised to add interference of the incoming speech with the respective response planning processes (Indefrey & Levelt, 2004; Levelt, 1989, 1992). With these processing pressures standing against the time pressures that are applied by the turn-taking system, there might be a level of processing at which the costs of early response planning match its benefits, the question being where that level is. One hypothesis is that only conceptual planning is done in overlap while formulation is postponed until the end of the incoming turn in order to avoid increased planning effort due to phonological interference. A competing hypothesis is that formulation, including word form retrieval, is done as early as possible and in overlap with the incoming turn in order to keep inter-turn gaps short.

These hypotheses are evaluated here in three experiments making use of a switch task. In Experiment 1, participants were required to name a presented object as fast as possible as a base task. In switch trials (25%), the object disappeared after having been presented for a short amount of time and was replaced by a word that had to be judged to be a real Dutch word or not by giving a button press response. These words were presented either after associated or phonologically related pictures or after unrelated pictures. Words that are associated to pictures are words that come to mind when a particular picture is presented, e.g. *cheese* when a mouse is presented. Phonologically related words on the other hand sound like the presented picture's name, e.g. *mouth* when a mouse is presented. In cases when the respective level of representation (lemma for associative relatedness or word form for phonological relatedness) was activated by the time of the task switch, relatedness of the target picture and the word replacing it should have an effect on participants' lexical decision performance. Assuming a structured mental lexicon consisting of at least three distinct levels of entries, namely concepts, grammatical or semantic entries (lemmas), and word-forms, to produce a word requires selecting the correct word form that belongs to the lemma matching the concept that should be expressed (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). To comprehend a presented word, on the other hand, requires the selection of a concept that belongs to a lemma matching the word form that was presented (Cutler, 2012; Norris, Cutler, McQueen, & Butterfield, 2006). For reading written words, a second word form representation, the orthographic representation, is assumed next to the phonological representation, with the two types of representation being linked in the lexicon, so that an activated orthographic representation leads to activation of the corresponding phonological representation (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Ellis & Young, 1988). If, at the time of the task switch, participants activated the lemma corresponding to the picture's name, association of the picture and the word replacing it should lead to associative facilitation (Alario, Segui, & Ferrand, 2000; La Heij, Dirkx, & Kramer, 1990; Perea & Rosa, 2002; Plaut, 1995). Similarly, if participants activated the word form of the picture to be named by the time of the task switch, the representations of phonologically related words should be suppressed below their level of resting activation, leading to decreased lexical decision performance in phonologically related words (Levelt et al., 1991; Pylkkänen, Gonnerman, Stringfellow, & Marantz, n.d.). As the processes of lemma selection are known to precede the processes of word form retrieval, three different stimulus-onset asynchronies (SOAs) are used in order to target the different levels of production planning (Levelt, 1989; Levelt et al., 1999).

Experiment 2 uses the same materials as Experiment 1 and takes an intermediate step between the monologic setup of Experiment 1 and the dialogic setup of

Experiment 3 by adding incoming questions being played to participants as distracting speech which participants were instructed to ignore. In that way, Experiment 2 will allow us to evaluate whether distracting speech as it is commonly used in experimental setups affects the timing of language planning.

In Experiment 3, the same materials were used as in Experiment 2. In Experiment 3 however, participants had to decide based on the question which one out of four displayed pictures they would have to name. In that way, the given task resembled a dialogical situation as participants were required to attend to the presented questions and answer them by naming one of the pictures. The format of these questions was designed to give away the cue to the target picture either during the middle of the question or only at its end. Again, in critical trials (25%), participants had to switch from the picture naming task to the lexical decision task. The relatedness effects of target pictures and words for lexical decision will shed light on the progress of response planning during the incoming turn on the one hand, as compared to at the end of the incoming turn on the other hand. Following the hypothesis that all stages of response planning are run through in overlap with the incoming turn (Levinson & Torreira, 2015) and consequently activating the respective representations on all levels of the mental lexicon, relatedness of the picture to be named and the word replacing it for lexical decision should have an effect on lexical decision performance both during the incoming question as well as at the end of it. If a relatedness effect was only found at the end of questions, however, when response planning is done in silence, and was absent in the middle of questions, where response planning is done in overlap, that finding would be taken as evidence for delayed response formulation. The filler trials (75%), in which participants have to overtly answer the question by naming the target picture, serve as a replication of the effects of planning in overlap that were described in the previous literature (Barthel et al., 2017, 2016; Bögels, Magyari, et al., 2015). If the responses are planned as early as possible, naming latencies should be faster in questions that give away the answer early as compared to questions that give away the answer only at their end.

2. Experiment 1

2.1. Method

Participants

Sixty-four Dutch native speakers were recruited as paid participants at Radboud University campus. Data of one participant was lost after recording. All participants reported to have normal or corrected to-normal vision and hearing as well as no speech or language impairments.

Apparatus

Participants were seated in a sound proof booth approximately 60 cm away from a 21 inch computer screen and a Sennheiser ME64 microphone. They were equipped with a two-precision-buttons response box based on USB-mouse script with 125 Hz sampling rate. Stimuli were presented using SMI ExperimentCenter software.

Materials and Design

256 pictures of objects were used in the experiment. The pictures were sourced online and are under the creative commons license. They were selected to be easy to recognise and name. 192 of these pictures served as filler objects in naming trials and were not systematically related to the pictures used in critical trials. The common names for these filler objects cover a broad range of medium frequency counts as extracted from the SUBLEX_NL corpus (Keuleers, Brysbaert, & New, 2010, mean log frequency per million = 1.95; SD = 1.8) and vary in length between one and five syllables (mean number of syllables = 2.4; SD = 0.95). The remaining 64 pictures served as critical objects in switch trials. The critical objects had very high name agreement, as assessed in a pretest with a different group of 31 participants (mean agreement = 96%, SD = 4%).

256 words were used in the lexical decision task, with half of them being real Dutch words (critical), the other half being pseudowords (filler). Each of the words was either associated with a critical picture or phonologically related to a critical picture's name (Type of Relation: associative/phonological), and would either be presented after the related picture or after another, unrelated picture (Relatedness: unrelated/related). Table 1 gives an overview of the tested conditions. Associatively related words were drawn from the Dutch Word Association Database (http://www.kuleuven.be/semlab/interface/index.php; see De Deyne & Storms, 2008), and were chosen to be strong associates of the picture name (mean first association strength = 30%, SD = 16%). Phonologically related words had the same syllable length and syllable structure as the related picture name and tended to differ from the picture name in one phoneme towards the end of the word (i.e. in nucleus, coda, or second syllable). Associatively related words were not phonologically related to the respective picture names, with maximally one overlapping segment (*mean overlap* = 5% of segments, SD = 11%). Phonologically related words were not associated with the pictures. Pseudoword

Table 1. Example item showing the four critical conditions tested in Experiment 1.

Type of relation	Relatedness	Target (name)	Lexical decision word
phonological	related	 (*appel*)	*ampel* (traffic light)
	unrelated	 (*zaag*)	*ampel* (traffic light)
association	related	 (*appel*)	*fruit* (fruit)
	unrelated	 (*zaag*)	*fruit* (fruit)

Each condition of an item was tested in a separate list.

strings were produced by changing one segment of one of the real words.

Eight experimental lists were constructed, with a different word following a given critical picture in each of the lists. Each participant was tested in one of the lists and assigned to one of three SOA groups (see Section *Procedure* below).

Procedure

Each trial began with a fixation cross in the middle of the screen for 2 s, followed by one of the pictures presented at the centre of the screen (see Figure 1). Participants were instructed to name the picture as fast as possible. The picture disappeared upon voice onset and was replaced by a blank screen for 1 second before the next trial started. In switch trials (25%), the picture was only presented for a short amount of time (SOA) before it was replaced by a letter string. Three SOA conditions of 70 ms, 120 ms, and 200 ms were tested between participants. Participants were instructed to abandon the naming task in case the picture was replaced by a word. In this case, participants were to decide whether the presented word was a real Dutch word or not, and give their response by pressing one of two buttons as fast as possible (with the "word" response lying on the right button). Upon pressing a button, the word disappeared and was replaced by a blank screen for 1 second before the next trial started. Every sixty-four trials, a pause screen was presented, giving participants the chance to take a short break.

The experiment proper was preceded by eight practice trials and followed by a post test in which participants were shown the 64 critical pictures and asked to name them, so as to check whether their responses matched the expected names for the critical pictures. The whole experimental session took about 40 min.

2.2. Results

Of the 12,096 naming trials, 481 trials (3.9%) were regarded as erroneous and consequently discarded, as the voice key was triggered more than four seconds after picture onset. Another 404 trials (3.4%) were discarded because they were outliers of more than 2.5 standard deviations by subject. Remaining trials had a mean naming latency of 1184 ms (SD = 488 ms; CI = ⟨1175 ms, 1193 ms⟩). Figure 2 shows a density plot of the distribution of naming latencies. Figures 1 and 2 in the Supplementary Materials show density plots of the distributions of naming latencies by SOA condition and by subject.

Of the 2016 critical lexical decisions, 111 (5.6%) were discarded since participants did not name the corresponding critical pictures by their standard labels in the post test. Inspecting the distributions of lexical decision latencies for each of the subjects, reaction times by two participants (both in SOA120 condition) were found to behave differently than those of the other subjects in not being uni-modally distributed, possibly hinting at the use of a reaction strategy ignoring

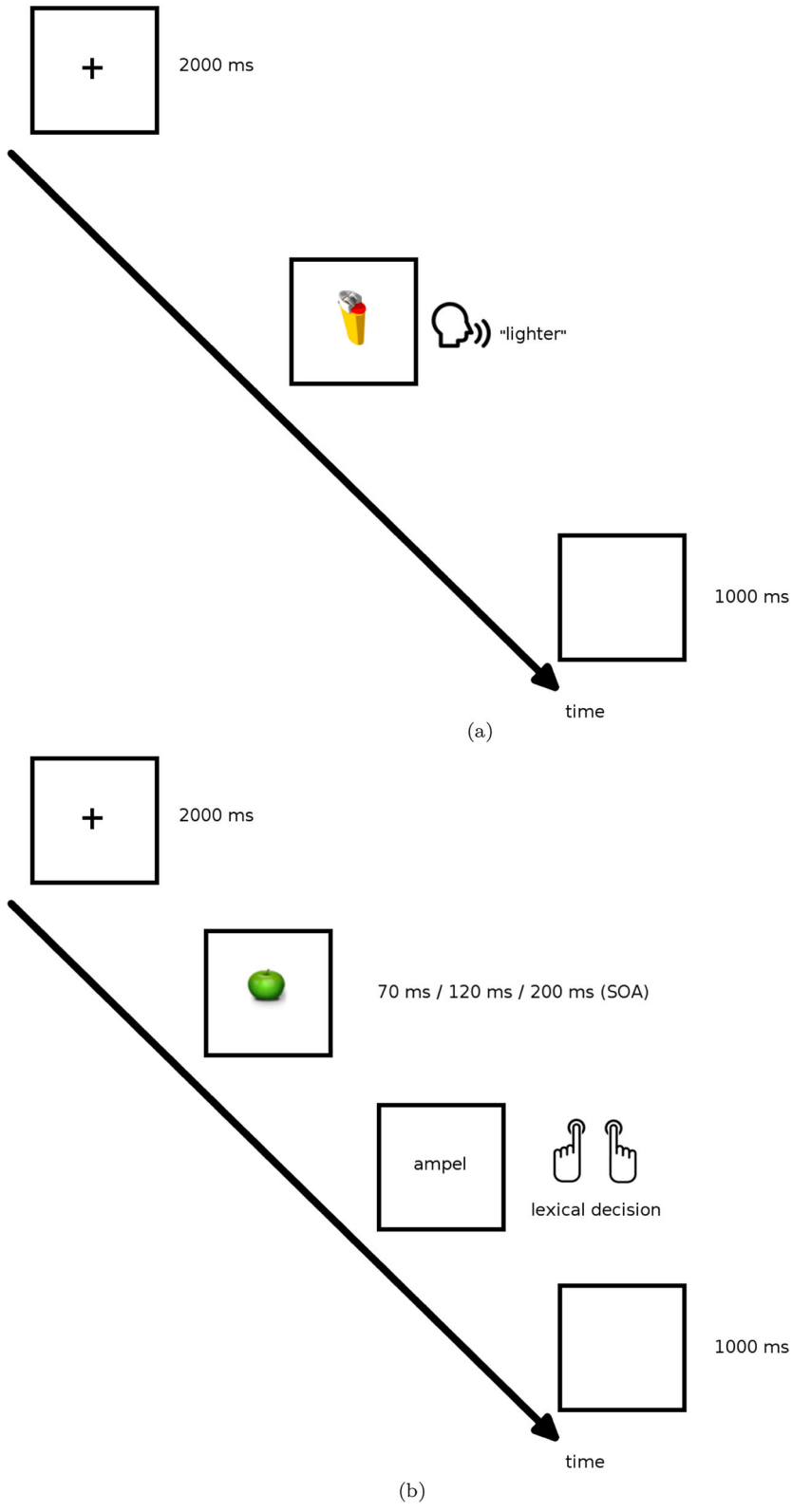

Figure 1. Timelines of a naming trial and a switch trial in Experiment 1. (a) naming trial and (b) switch trial.

the instruction to give a decision as fast as possible. Data from these two subjects were excluded from further analyses.[2] 246 button press responses (13.3%) were erroneous. Notably, almost twice as many errors were produced with words that were presented after phonologically related pictures (28.2%) as compared to after phonologically unrelated pictures (15.1%, see Figure 3).

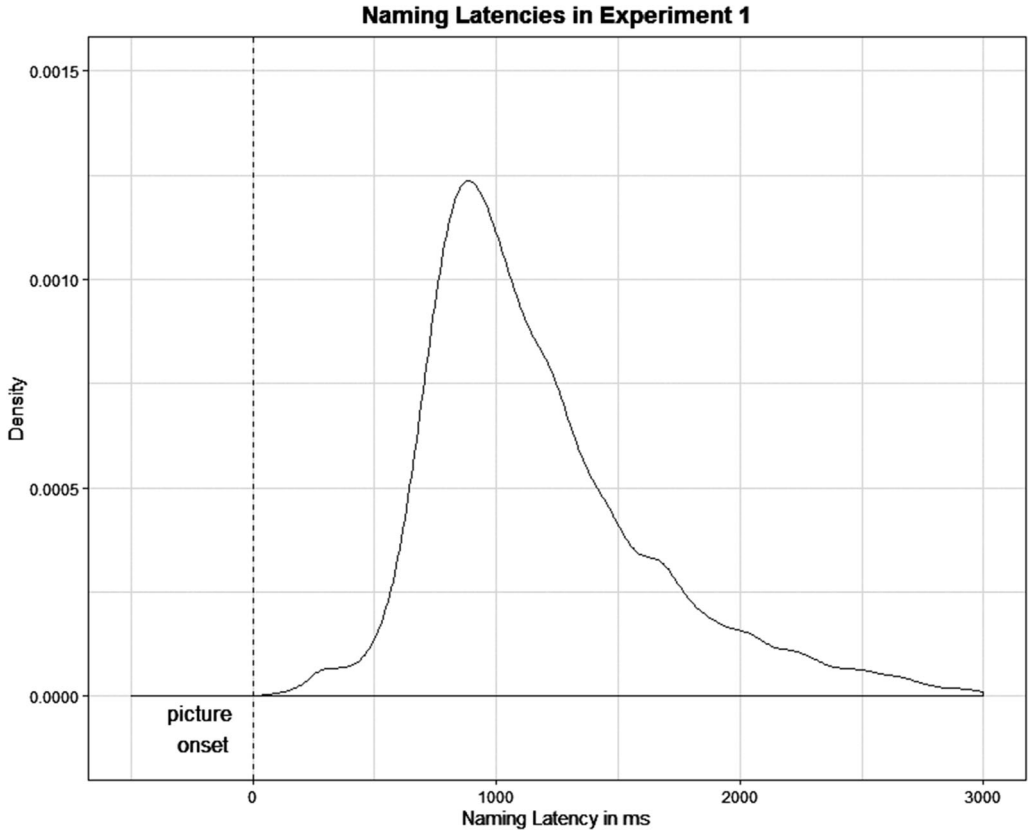

Figure 2. Distribution of naming latencies in Experiment 1.

Statistical analyses have been conducted with R (R Core Team, 2019). Mixed effects regression models have been fitted using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) and predictors' statistical significance was assessed with *F*-tests with Kenward-Roger approximations of degrees of freedom (Fox & Weisberg, 2011; Halekoh & Hojsgaard, 2014; Kenward & Roger, 1997). Bayesian liner models have been fitted using the brms package (Bürkner, 2017) and 3000 iterations. Bayes factors were calculated using the built in brms *hypothesis*-function.[3] Throughout the study's experiments, the maximal random effects structures justified by design which allowed models to converge were used (Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013), with subject and item as random effects. All categorical predictors were deviation coded with the exception of SOA in Experiment 1, which was simple coded with the intercept referring to the grand mean of the three levels of the factor and the effect of the first two levels (SOA70 and SOA120) being compared to the effect of the third level (SOA200).

Error rates were analysed in a logit mixed effects regression model with SOA, Relatedness and Type of Relation as well as their interactions as predictors (see Table 1 in Supplementary Materials). While mean error rates in SOA70 and SOA200 do not differ significantly, error rates in SOA120 are significantly higher than error rates in SOA200 ($\beta = 0.940$, SE $= 0.407$, $z = 2.307$, p.05). This effect, however, does not significantly interact with Relatedness nor with Type of Relation and is hence probably due to differences between the tested populations. The interaction effect between Relatedness and Type of Relation is significant ($\beta = 1.011$, SE $= 0.467$, $z = 2.164$, p.05), indicating that the main effect of Relatedness differs between the phonological and the associative sets of words. To further investigate the effect of Relatedness, corrected post-hoc tests based on estimated marginal means have been calculated using the emmeans package (Lenth, 2019). Relatedness was significant in phonologically related words ($F = 21.269$, $p < .001$), but not in associatively related words ($F = 0.020$, $p = .88$), indicating that participants made more errors when words were presented after phonologically related pictures than after non-related pictures and that error rates did not differ between words that were presented after associated pictures versus after non-related pictures.

Erroneous trials were discarded from the following analyses of lexical decision latencies. Further, 39 (2.4%) trials were discarded because their reaction latencies were outliers of more than 2.5 standard deviations by subject. The mean button press latency of the remaining 1560 trials was 1118 ms (SD $=$ 300 ms, see Figure 3).

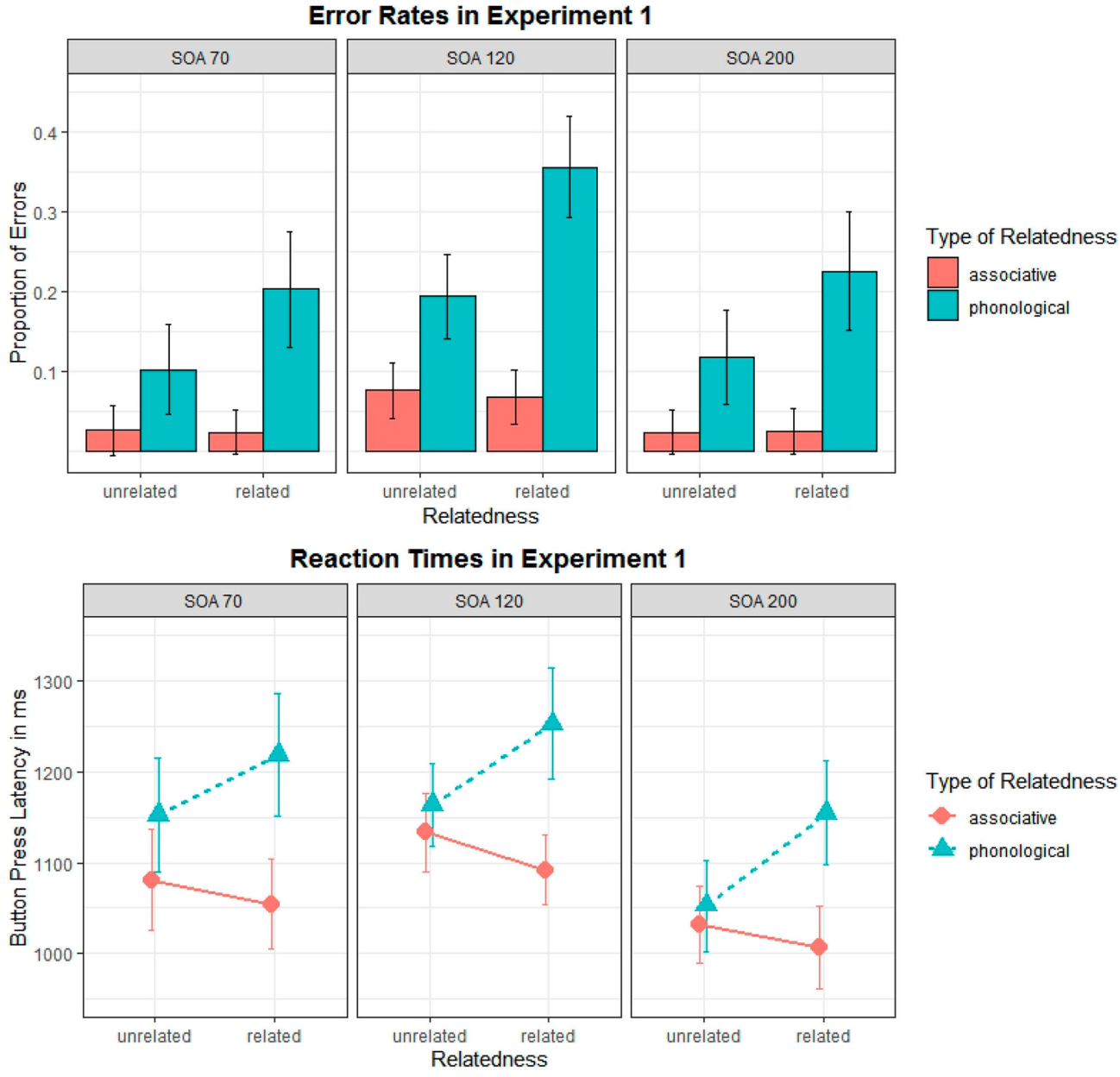

Figure 3. Reaction times and error rates in lexical decisions in Experiment 1. Bars represent 95% confidence intervals.

The log-transformed button press latencies of correct trials were analysed in a linear mixed effects regression model with SOA, Relatedness and Type of Relation as well as their interactions as predictors (see Table 2 in Supplementary Materials). The interaction effect of Relatedness × Type of Relation turned out to be highly significant ($\beta = 0.040$, SE = 0.008, $t = 4.82$, $F = 23.196$, df = 1422, p.001), indicating that the effect of Relatedness goes in opposite directions in the phonological and associative sets of words. To further investigate the effects of Relatedness, corrected post-hoc tests based on estimated marginal means have been calculated. In these tests, Relatedness turns out to significantly affect decision times in both associative words and phonological words with the effect going in opposite directions. While decisions in the associative set were made faster when the words were presented after associated pictures than after unrelated pictures ($\beta = -0.012$, SE = 0.005, $p < .05$), decisions in the phonological set were made slower when the words were presented after phonologically related pictures than after non-related pictures ($\beta = 0.028$, SE = 0.006, $p < .001$). To test which level of SOA showed the most robust effects of Relatedness, corrected post-hoc tests based on estimated marginal means have been calculated. While none of the effects of association survived the correction for multiple comparisons, the effect was still marginally significant in SOA120 (SOA70: $\beta = 0.012$, SE = 0.010, $p = .250$; SOA120: $\beta = 0.014$, SE = 0.007, $p = .061$;

Table 2. Bayesian linear regression model on button press latencies in Experiment 1.

	β	SE	lower CrI	upper CrI
Intercept	1118.96	27.60	1065.66	1174.18
SOA70	56.83	71.71	−81.29	200.81
SOA120	99.75	64.85	−26.84	228.27
Relatedness	22.29	14.46	−6.91	50.48
Type of Relation	98.12	13.34	71.88	124.35
SOA70 × Relatedness	−21.18	34.58	−88.54	45.03
SOA120 × Relatedness	−20.92	31.71	−82.33	41.22
SOA70 × Type of Relation	17.05	33.35	−47.96	82.72
SOA120 × Type of Relation	5.25	32.57	−56.45	69.68
Relatedness × Type of Relation	107.47	28.92	49.66	164.28
SOA70 × Rel. × Type of Rel.	−17.98	63.91	−142.18	107.94
SOA120 × Rel. × Type of Rel.	3.90	57.31	−107.48	114.01

For comparison of the presented effects of SOA, SOA200 was used as a baseline. Credible intervals contain 95% area under the posterior likelihood distribution. Model formula = Latency ~ intercept + SOA * relatedness * type.of.relation + (intercept + SOA * relatedness * type.of.relation | subject) + (intercept + SOA * relatedness * type.of.relation | item).

SOA200: $\beta = 0.008$, SE = 0.010, $p = .395$). The effect of phonological relatedness was significant in all three levels of SOA and turned out to be most pronounced in SOA200 (SOA70: $\beta = -0.022$, SE = 0.011, p=.044; SOA120: $\beta = -0.024$, SE = 0.009, $p = 0.006$; SOA200: $\beta = -0.037$, SE = 0.011, $p = .001$).

In order to test for the likelihood distribution of the obtained reaction times effects, a Bayesian linear model was used to fit decision latencies, with Relatedness, Type of Relation and SOA as well as their interactions as predictors with default uninformative priors and maximal random effects structures for both subjects and items (see Table 2). If 0 lies outside the credible interval, there is sufficient evidence to suggest there is an effect of a particular predictor. As the effect of Relatedness turned out to be decisively affected by the Type of Relatedness ($\beta = 107$ ms, SE = 29 ms, CrI = ⟨50 ms, 164 ms⟩, BF = inf), we conducted two Bayesian inference tests testing the effects of Relatedness separately for the associative and phonological sets of words. The first test revealed decisive evidence for the effect of Relatedness in the associative set of words, with decisions for words being faster when they are presented after associated pictures than when they are presented after non-related pictures ($\beta = 31$ ms, SE = 18 ms, CrI = ⟨1 ms, 62 ms⟩, BF = 23). The second test revealed decisive evidence for the effect of Relatedness in the phonological set of words, with decisions for words being slower when they are presented after phonologically related pictures than when they are presented after non-related pictures ($\beta = -76$ ms, SE = 22 ms, CrI = ⟨−113 ms, − 39 ms⟩, BF = 1999).

2.3. Discussion

Experiment 1 examined the effects of associative relatedness and phonological relatedness of pictures and words on lexical decision performance in a switch task. Participants were instructed to name displayed pictures as fast as possible as a base task. In 25% of trials, the picture was replaced by a word without prior notice after 70 ms, 120 ms, or 200 ms (SOA) and participants had to abandon the naming task and give a lexical decision response instead, evaluating whether the word was a real Dutch word or not. Decisions were faster if words were presented after pictures that were associated with the words than after pictures that were unrelated to the words, and decisions were slower and yielded more errors if words were presented after a picture whose name was phonologically related to the word as compared to when they were presented after an unrelated picture, with this effect of phonological inhibition being most pronounced at an SOA of 200 ms. The effect of associative facilitation was weaker in absolute terms then the effect of phonological inhibition and only showed in participants' reaction times but not in their error rates. One possible reason for the effects of associative relatedness being weaker than the effects of phonological relatedness might be that association strengths between target pictures and words might vary greatly between participants or were generally too low across participants for activation to spread reliably to the lemmas of the lexical decision words while participants prepared to name the picture. Moreover, Jongman and Meyer (2017) found that effects of associative relatedness disappear in situations where task switches are unpredictable. In their picture naming study, associated auditory primes affected naming latencies only when the task was held constant across trials but did not affect latencies when task switches were unpredictable (as was the case in the present study). Nonetheless, since effects of phonological relatedness were observed reliably throughout Experiment 1, semantic processing of the pictures must have taken place by the time of the respective SOA's. Consequently, we will drop the associative condition and focus on phonological relatedness in Experiment 2, where we aim to replicate the results of phonological inhibition obtained in Experiment 1 in the presence of distracting incoming speech. As the target effect was most robust at SOA 200, we will focus on that SOA in the following Experiment.

3. Experiment 2

3.1. Introduction

In Experiment 2 we take an intermediate step towards a dialogic test situation by adding background speech to the switch task used in Experiment 1. While participants dealt with the respective tasks (picture naming as base

task; lexical decision as switch task in 25% of trials), they were auditorily presented with one question per trial in order to test whether the same effects of phonological inhibition can be observed at the same SOA as in Experiment 1 if participants are presented with distracting speech input while attending to the switch task. If so, the same SOA can be used in a question-answer task in Experiment 3. If not, one probable reason this test might fail to replicate the previous results is that the speech production processes involved in the picture naming task get delayed or slowed down by distracting speech. In that case, the SOA to be used in Experiment 3 should be longer than in Experiment 2.

Based on the results obtained in Experiment 1, Experiment 2 focuses on effects of phonological relatedness of picture names and words for lexical decision at an SOA of 200 ms.

3.2. Method

Participants

Sixteen Dutch native speakers who did not take part in Experiment 1 were recruited as paid participants on Radboud University campus. All participants reported to have normal or corrected to-normal vision and hearing as well as no speech or language impairments.

Apparatus

The apparatus was the same as in Experiment 1, except that participants were additionally equipped with closed headphones.

Materials and Design

The materials used in in Experiment 1 were also used in Experiment 2. Additionally, 256 questions that had been pre-recorded by a male speaker were used. Each question asked for one of the pictures used in the experiment. Questions were of the format "Which object that has property X also has property Y?" Example: "Which object that grows on a tree is also edible?" The 64 questions that were used in switch trials were also used in a second version with the mentioned properties in a swapped order ("Which object that is edible also grows on a tree?") (Question Type: A/B). The same questions in these two types will also be used in Experiment 3, where the questions are relevant to the task of participants and give away their answer either early or late. For now, however, participants were instructed to ignore the questions. Questions had a mean length of 3.74 s (SD = 0.39 s).

Eight experimental lists were constructed, with a different word following a given critical picture in half of the lists, while a question of either type was played. The same words followed a given critical picture in the other half of the lists, while a question of the other type was played. Each participant was tested in one of the lists.

Procedure

Each trial began with a fixation cross in the centre of the screen while a question was played. Participants were instructed to completely ignore the questions. In the middle of the question, at the beginning of the phrase stating either the first (Question Type A) or the second property that was mentioned in the question (Question Type B), the picture corresponding to the question would replace the fixation cross and participants were instructed to name the picture as fast as possible. The picture disappeared upon voice onset and was replaced by a blank screen for 1 second before the next trial started. In lexical decision trials, the picture was replaced by a word after being presented for 200 ms (SOA). In these critical trials, participants were instructed to abandon the naming task and instead press one of two buttons indicating whether the word was a real Dutch word or not. Upon pressing a button, the word disappeared and was replaced by a blank screen for 1 second before the next trial started. Every sixty-four trials, a pause screen was presented, giving participants the chance to take a short break.

The experiment proper was preceded by eight practice trials and followed by a post test in which participants were shown the sixty-four critical pictures and asked to name them, so as to check whether their responses matched the expected names for the critical pictures. The whole experimental session lasted about 50 min.

3.3. Results

Inspecting the distribution of naming latencies for each subject, naming latencies of one subject were found to differ from those of the other subjects in being bimodally distributed, possibly indicating the use of a waiting strategy that diverges from normal production planning. Data of that subject were removed from analyses.[4] Of the remaining 2880 naming trials, 120 trials (4.2%) were regarded as erroneous and consequently discarded, as the voice key was triggered more than four seconds after picture onset. Another 101 trials (3.7%) were discarded because they were outliers of more than 2.5 standard deviations by subject. Remaining trials had a mean naming latency of 1225 ms (SD = 588 ms; CI = ⟨1202, 1247⟩; Figure 4). The log-transformed naming latencies were analysed in a mixed effects model with Question Type as predictor. Naming latencies did not significantly differ between the two levels of

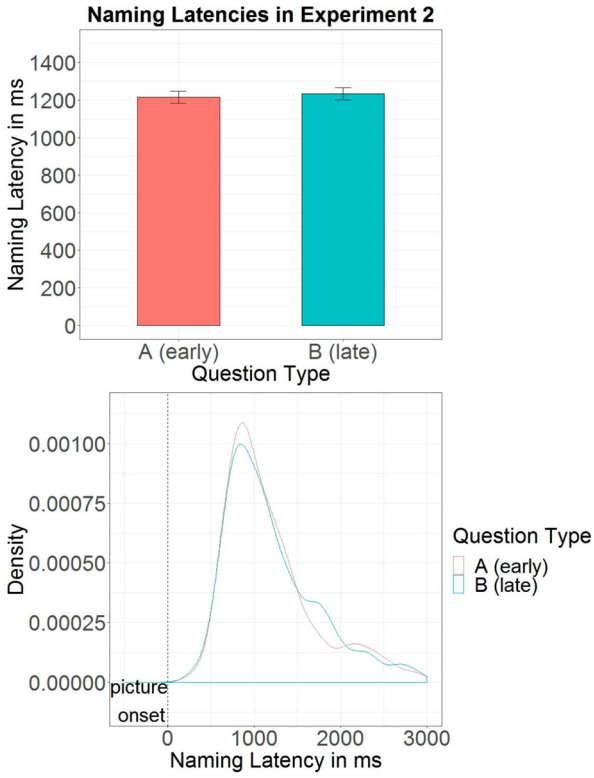

Figure 4. Naming latencies in Experiment 2. Bars represent 95% confidence intervals.

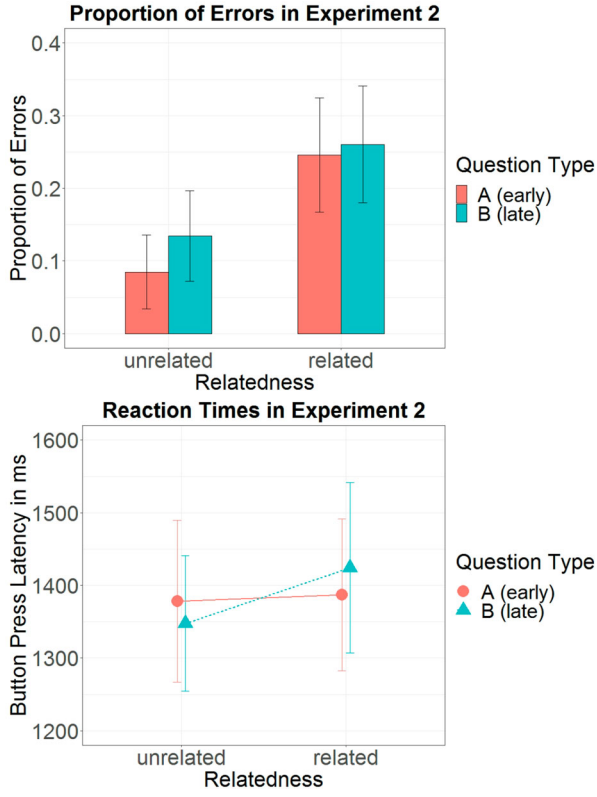

Figure 5. Reaction times and error rates in lexical decisions in Experiment 2. Bars represent 95% confidence intervals.

Question Type ($\beta = 0.007$, SE = 0.006, $t = 1.22$, $F = 1.492$, $p = .221$). An independent t-test comparing naming performance in Experiments 1 and 2 shows naming latencies to be significantly longer in Experiment 2 ($t = -3.31$, df = 3578, $p < .001$, CI = ⟨17 ms, 65 ms⟩).

Of the 480 critical lexical decisions, 6 (1.3%) were discarded since participants did not name the corresponding critical pictures by their standard labels in the post test. Another 86 button press responses (18.1%) were erroneous (see Figure 5). Notably, more than twice as many errors were produced when words were presented after pictures with related (25.3%) as compared to unrelated names (11%).

Error rates were analysed in a logit mixed effects regression model with Relatedness and Question Type as well as their interaction as predictors (see Table 3 in Supplementary Materials). Relatedness significantly affected error rates, with participants making more errors in related than in unrelated words ($\beta = 1.404$, SE = 0.407, $z = 3.449$, $p < .001$). The main effect of Question Type as well as its interaction with Relatedness turned out non-significant.

Erroneous trials were discarded for the following analyses of decision latencies. Moreover, 10 (2.6%) trials were discarded because their reaction latencies were outliers of more than 2.5 standard deviations by subject. The mean button press latency of the remaining 378 correct trials was 1382 ms (SD = 520 ms). See Figure 5 for decision latencies and error rates by condition.

Button press latencies of correct trials were analysed in a mixed effects model with Relatedness and Question Type as well as their interaction as predictors (see Table 4 in Supplementary Materials). The main effect of Relatedness was not significant ($\beta = 0.015$, SE = 0.014, $F = 1.904$, p=.302), neither was there a significant interaction of Relatedness with Question Type ($\beta = 0.013$, SE = 0.019, $F = 0.388$, $p = .483$). The main effect of Question Type was also non-significant ($\beta = 0.009$, SE = 0.009, $F = 0.002$, $p = .345$).

To test for the reliability of the attested null results and to get an estimation of the distribution of probability of the observed relatedness effect, a Bayesian linear model was used to fit decision latencies, with Relatedness and Question Type as well as their interaction as predictors and maximal random effects structures for both subjects and items (see Table 5 in Supplementary Materials). A normal prior distribution for the expected effect of Relatedness was used, with the mean being the mean Relatedness effect observed in Experiment 1 (74 ms) and the tenfold standard deviation of that previously observed effect (250 ms), so as to make the prior moderately informative. A Bayesian inference test

testing for the modelled effect of Relatedness yielded very weak evidence for the effect being higher than zero ($\beta = 51$ my, SE $= 57$ ms, CrI $= \langle -40$ ms, 148 ms$\rangle$, BF $= 4.68$). Similarly, a second test for the the modelled interaction effect of Relatedness × Question Type yielded very weak evidence for the effect being higher than zero ($\beta = 71$ my, SE $= 82$ ms, CrI $= \langle -65$ ms, 203 ms$\rangle$, BF $= 4.27$).

3.4. Discussion

After phonological relatedness of picture names and words for lexical decision led to interference effects in response latencies and error rates in Experiment 1, the present Experiment was designed to replicate these results with distracting background speech. In particular, it was run to test whether the same SOA of 200 ms that led to phonological interference in Experiment 1 can be expected to yield comparable interference effects in the presence of distracting speech or whether response planning gets slightly delayed or slowed down.

Even though participants seemed to ignore the incoming questions, as evidenced by very similar naming latencies in both question types, the effects on lexical decision performance obtained in Experiment 1 could not be fully replicated. While the significant effect of Relatedness on error rates indicates phonological interference in both early and late questions, Relatedness did not have a reliable effect on reaction times. As the results of the Bayesian analyses have not yielded decisive evidence for the presence or absence of a Relatedness effect, it is possible that a potential effect could not have been detected due to a lack of statistical power. It therefore remains unclear whether participants were planning their verbal responses phonologically during the incoming questions or not. However, as the naming latencies obtained in naming trials were on average 41 ms longer than in Experiment 1, it is likely that incoming speech slowed down the processes of response planning. That means that the results of Experiment 1 (with an SOA of 200 ms) might not fully replicate in the question-answer situation we aim to test in Experiment 3. For that reason, Experiment 3 was designed to use a longer SOA of 300 ms.

4. Experiment 3

4.1. Introduction

Following Experiment 2, in which questions were presented as distracting background speech that had to be ignored by participants, Experiment 3 makes use of a dialogic task in order to investigate whether next speakers plan their utterance phonologically in overlap with the incoming turn. Participants have to attend to auditorily presented questions in order to be able to answer them. The questions ask for one out of four pictures of objects that are presented to participants. They are designed so that they give away their answer either in the middle of the question or only at their end. In that way, speech planning in overlap, which is expected in questions giving away the answer early, can be compared to speech planning in silence, which is expected after questions giving away the answer at their end. Participants' eye-gaze is used as an indicator for the initiation of response planning, assuming that speakers fixate the object they mentally process at a given moment (Barthel et al., 2016; Just & Carpenter, 1980; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In a quarter of trials, the task switches after 300 ms of gaze falling on the target object and participants have to give a lexical decision instead of answering the question. The relatedness of the lexical decision words and the target picture was manipulated in order to investigate whether participants already retrieved the word form of the picture name in overlap with the incoming question or not. If phonological planning was delayed until the end of the incoming turn, relatedness of the lexical decision word and the target picture name should have no effect on lexical decision performance. If, on the other hand, participants planned their answer phonologically already during the incoming question, phonological relatedness of the lexical decision word and the picture name should affect lexical decision performance. If the word form of the picture name was already retrieved by the time of the task switch, activation of the lexical decision word should be inhibited, leading to longer decision latencies and increased error rates.

4.2. Method

Participants

Forty-five Dutch native speakers who did not take part in Experiments 1 and 2 were recruited as paid participants on Radboud University campus. All participants reported to have normal or corrected to-normal vision and hearing as well as no speech or language impairments. In thirteen participants tested in a first test session, more than 25% of the critical lexical decision trials were invalid, either due to trackloss of participants' gaze or because participants kept on naming the target picture even though they were instructed to abandon the naming task and switch to lexical decision. Data of these participants were discarded. The other thirty-two participants were tested in a second session on a second experimental list (see Section *Materials*

and Design below), with at least one day between the two test sessions. In one of these participants, more than 25% of the critical trials of the second test session were invalid. This participant was replaced.

Apparatus

The apparatus was the same as in Experiment 2, except that participants' eye-movements were monitored with an SMI RED-m remote eye-tracker (120 Hz sampling rate).

Materials and design

The same 256 pictures of objects that were used in Experiments 1 and 2 were used as target pictures in Experiment 3, 192 in naming items (75%) and 64 in lexical decision items (25%). In each naming item, four pictures were displayed in the four corners of the screen, with white space between each of the pictures. Each picture was used as target in one naming item and served as a distractor picture in another three naming items. Similarly, in each lexical decision item, four pictures were displayed in the four corners of the screen, with one of the 64 critical pictures as the target picture in one of the display's corners. 192 additional pictures that had not been used in the previous Experiments were used as distractor pictures in the 64 lexical decision items, so that each critical picture would only be displayed once per test session. The position of the target picture on the screen was balanced across the experiment.

256 questions that were used in Experiment 2 were used in Experiment 3, each question asking for one of the target pictures. The questions were of the format "Which object that has property X also has property Y?" One of these properties was uninformative, as all four pictures on the display (target and distractors) carried that property. The other property was informative, since only the target picture carried that property. In lexical decision items, two versions of the respective question were used, with the order of the properties mentioned in the question being swapped between the two versions. The informative property was therefore available either early or late during the question (Question Type: early/late), as illustrated in the following example of a lexical decision trial with the pictures of an apple, a potato, a strawberry, and a broccoli, playing either the question "Which object that grows on a tree is also edible?" (Question Type: early) or the question "Which object that is edible also grows on a tree?" (Question Type: late). Naming items were coupled with only one question, half of the naming items using early questions and the other half using late questions. The questions had a mean length of 3.74 s (SD = 0.39 s). See Table 8 in Supplementary Materials for a list of Materials.

In lexical decision trials (25%), the four pictures were replaced by a word appearing at the position of the target picture. The same words for lexical decision that were used in Experiments 1 and 2 were re-used in Experiment 3, with half of the words being real Dutch words, the other half being pseudowords. The words were either presented after target pictures whose name was phonologically related or after pictures whose name was unrelated (Relatedness: unrelated/related).

Eight experimental lists were constructed, with a different word following a given critical picture in half of the lists, while a question of either type was played. The same words followed a given critical picture in the other half of the lists, while a question of the other type was played. Each participant was tested in two of the lists, with at least one day between the two test sessions.

Procedure

Each trial began with a fixation cross for 1.5 s to attract participants' gaze to the centre of the screen (see Figure 6). The fixation cross was replaced by a display showing four pictures of objects in the four corners of the screen. The pictures were approximately 450 × 450 pixels large and occupied about 2.5° of participants' visual angle. 750 ms after the pictures appeared, a question was played. In naming trials (75%), participants had to answer the question as fast as possible by naming one of the displayed objects. Upon voice onset, the pictures would be replaced by a blank screen for 1.5 s before the next trial would start. In switch trials (25%), the pictures were replaced by a word that would appear in the position of the target object as soon as the participant's gaze would dwell on the target object for 300 ms (SOA), measured from the onset of the informative part of the particular question in a given trial. That part of the question began on average either after 1.34 s (in early questions; SD = 0.35 s) or after 2.96 s (in late questions; SD = 0.44 s). In these switch trials, participants were to abandon the naming task and switch to deciding whether the presented word was a real Dutch word or not and give their response by pressing one of two buttons as fast as possible (with the "word" response on the right button). Upon button press, the word would be replaced by a blank screen for 1.5 s before the next trial would start. Every sixty-four trials, a pause screen was presented, giving participants the chance to take a short break. At the beginning of the experiment, as well as after each of the short breaks, the eye-tracker was calibrated on nine points of the screen.

The experiment proper was preceded by eight practice trials and followed by a post test in which participants were shown the sixty-four critical pictures and asked to name them in order to check whether their

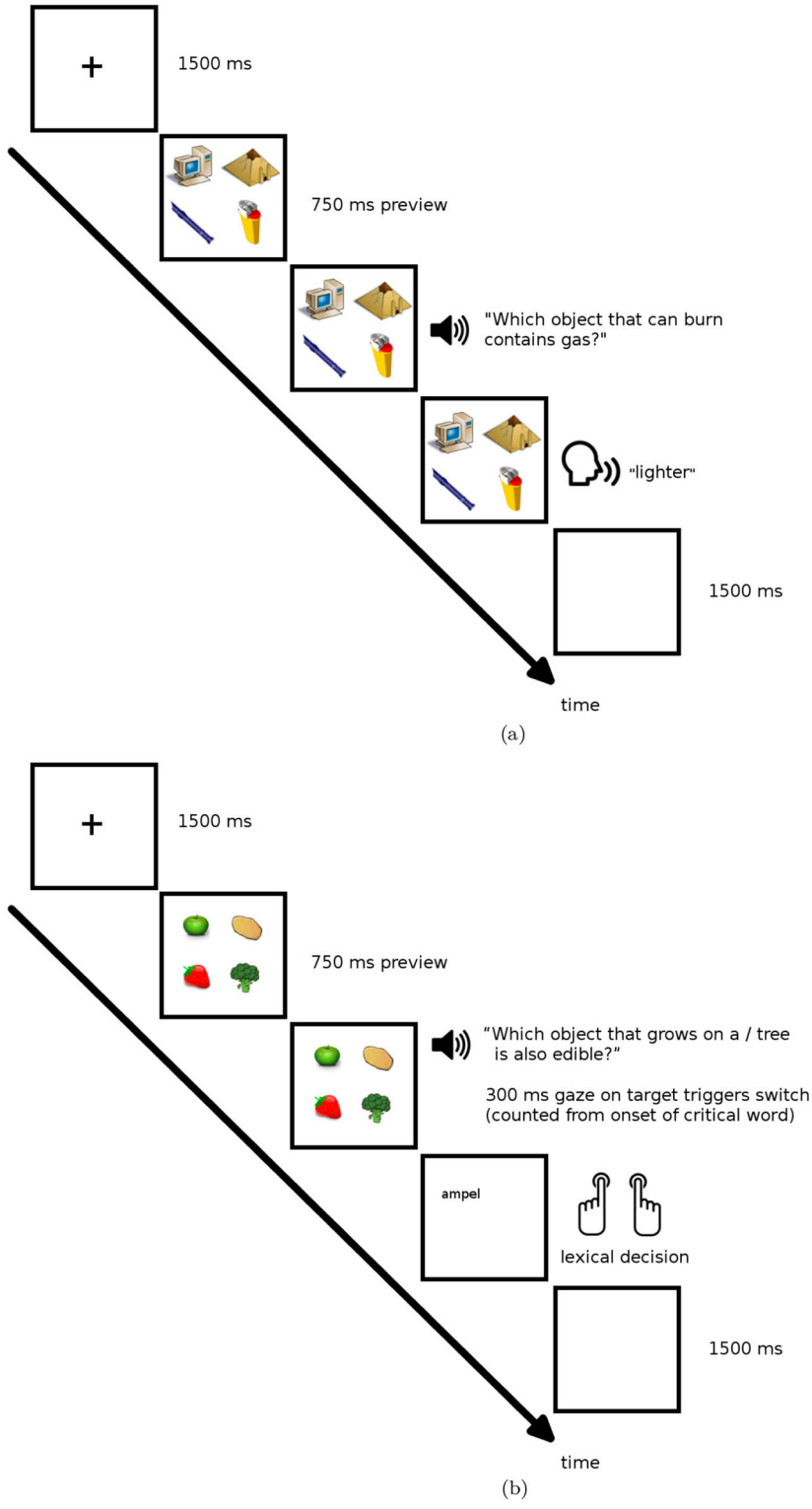

Figure 6. Timelines of trials in Exp. 3 exemplified with translations of early questions./ = beginning of the informative word in the question. Dutch originals: "Welk object dat kan branden, bevat gas?" in naming trial and "Welk object dat aan een boom groeit is ook eetbaar?" in switch trial. (a) naming trial and (b) switch trial.

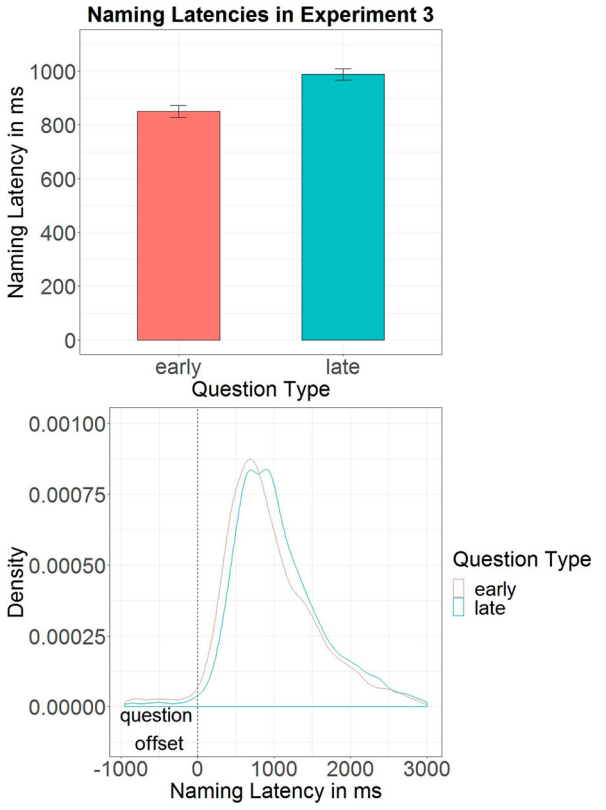

Figure 7. Naming latencies in Experiment 3. Bars signify 95% confidence intervals.

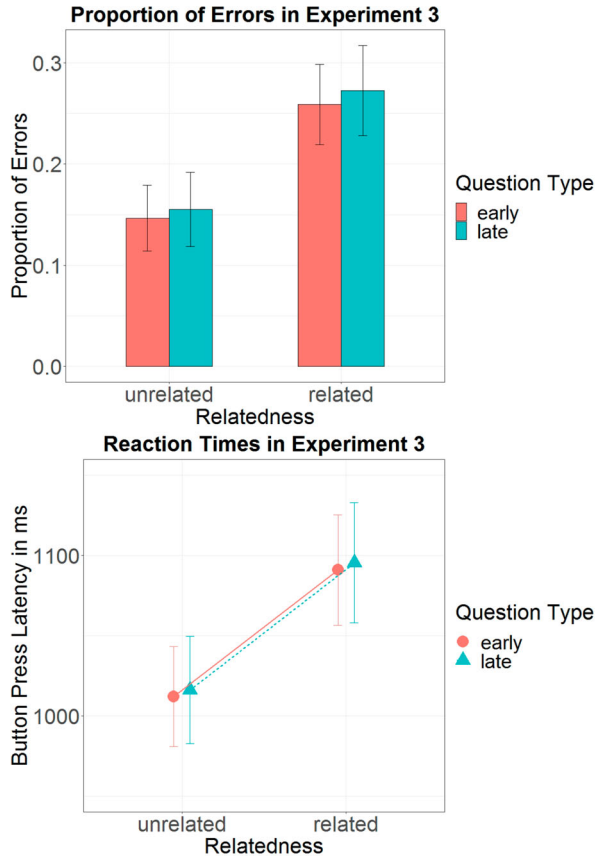

Figure 8. Reaction times and error rates in lexical decisions in Experiment 3. Bars signify 95% confidence intervals.

responses matched the expected names for the critical pictures. The whole experimental session lasted about 50 min.

4.3. Results

Inspecting the distribution of naming latencies for each subject, naming latencies of one subject were found to differ from those of the other subjects in being bimodally distributed, possibly indicating the use of a waiting strategy that diverges from normal production planning.[5] 390 naming trials (3.3%) were regarded as erroneous reactions as they triggered the voice key either more than two seconds before the end of the question (when the answer could not yet have been known) or more than four seconds after the end of the question and were consequently discarded. Another 179 trials (1.6%) were discarded because they were outliers of more than 2.5 standard deviations by subject and test session. The remaining 11,342 naming trials had a mean naming latency of 919 ms (SD = 848 ms; Figure 7), measured from the end of the question. A mixed effects model on log-transformed naming latencies with Question Type as predictor revealed a significant main effect of Question Type, with naming latencies being shorter in early question trials than in late question trials ($\beta = 0.006$, SE = 0.002, $F = 4.258$, df = 54, $p < .05$).

Of the 1984 critical lexical decision trials, 37 (1.8%) were discarded because participants' name for the respective target objects in the post test did not match the standard label for the object. Moreover, 84 (4.3%) trials were discarded due to trackloss of participants' gaze direction during the trial. 166 (8.9%) critical trials were discarded because participants overtly named at least part of the target picture, contrary to instructions. Of the remaining trials, 354 (20.8%) decisions were erroneous (see Figure 8). Error rates in related trials (26.5%) were 81% higher than in unrelated trials (15%).

Error rates were analysed in a logit mixed effects regression model with Relatedness and Question Type as well as their interaction and Test Session (1/2) as predictors (see Table 6 in Supplementary Materials). Participants made marginally significantly less errors in the second as compared to the first test session, indicating a practice effect between test sessions ($\beta = -0.333$, SE = 0.177, $z = -1.876$, $p = .06$). While the main effect of Question Type and its interaction with Relatedness are non-significant, the main effect of Relatedness

turns out significant ($\beta = 0.832$, SE $= 0.135$, $z = 6.134$, $p < .001$), with participants making more errors when words for lexical decision are presented after target pictures with related names than when they are presented after target pictures with unrelated names.

Erroneous trials were discarded for the following analyses of decision latencies. Furthermore, 33 (2.4%) trials were discarded because they were outliers of more than 2.5 standard deviations per subject and test session. The mean button press latency (RT) in the remaining 1311 correct lexical decision trials was 1051 ms (SD = 315 ms; Figure 8).

Participants triggered the change of display on average after 2417 ms in early questions and after 3888 ms in late questions. Given that the questions had a mean length of 3.74 s, displays were generally changed in overlap in early questions and in silence in late questions.

The log-transformed button press latencies were analysed in a mixed effects model with Relatedness, Question Type and Test Session (1/2) as well as the interaction of Relatedness and Question Type as predictors (see Table 7 in Supplementary Materials). Test Session significantly affects decision times, with participants taking faster decisions in the second test session than in the first test session, showing a training effect between sessions ($\beta = -0.068$, SE $= 0.008$, $F = 61.597$, df $= 30$, $p < .001$). The main effect of Relatedness turns out significant ($\beta = 0.028$, SE $= 0.005$, $F = 23.286$, df $= 42$, $p = .001$), with decisions being slower when words for lexical decision are presented after target pictures with related names than when they are presented after target pictures with unrelated names. Question Type ($\beta = 0.001$, SE $= 0.007$, $F = 0.046$, df $= 40$, p=.832) as well as its interaction with Relatedness ($\beta = -0.001$, SE $= 0.011$, $F = 0.012$, df $= 39$, $p = .912$) turn out non-significant in the model.

In order to test for the likelihood distribution of the obtained effect of Relatedness on reaction times and the evidence for the absence of an interaction effect of Relatedness × Question Type, a Bayesian linear model was used to fit decision latencies, with Relatedness and Question Type as well as their interaction and Test Session as predictors and maximal random effects structures for both subjects and items (see Table 3). Based on the obtained Relatedness effects in Experiments 1 and 2, we set a normally distributed prior with the mean of the previously observed effects (66 ms) and the tenfold SD of these effects (210 ms), in order to make the prior moderately informative. A Bayesian inference test based on the model revealed substantial evidence for the absence of an interaction effect of Relatedness × Question Type ($\beta = 1$ ms, SE $= 30$ ms, CrI $= \langle -58$ ms, 60 ms$\rangle$, $BF =$ 7.15), indicating that the effect of Relatedness does not differ between early and late questions. A second Bayesian inference test yielded decisive evidence for the observed main effect of Relatedness to be higher than zero ($\beta = 72$ ms, SE $= 16$ ms, CrI $= \langle 46$ ms, 98 ms$\rangle$, BF $=$ inf; see Figure 9), indicating that decision latencies were longer when words were presented after pictures with phonologically related names then when they were presented after unrelated pictures.

Table 3. Bayesian linear regression model on button press latencies in Experiment 3.

	β	SE	lower CrI	upper CrI
Intercept	1156.30	46.41	1063.81	1250.13
Test Session	−173.97	27.45	−227.47	−118.66
Relatedness	71.81	15.93	40.64	102.63
Question Type	2.87	22.99	−40.91	48.31
Rel. × Question Type	0.91	30.15	−57.71	60.01

Credible intervals contain 95% area under the posterior likelihood distribution. Model formula = brm(1 + Test.Session + Relatedness ∗ Question.Type + (1 + Test.Session + Relatedness ∗ Question.Type | subject) + (1 + Test.Session + Relatedness ∗ Question.Type | item)).

4.4. Discussion

Experiment 3 tested the time course of speech production planning in overlap with an incoming turn that requires a response. In each incoming turn, participants heard a question they had to answer by naming one of four pictures. These questions either gave away their answer already in the middle of the question or only at their end. Naming latencies were shorter when the answer to the question became clear earlier. This finding is taken to indicate that participants profited from planning in overlap with early questions, replicating

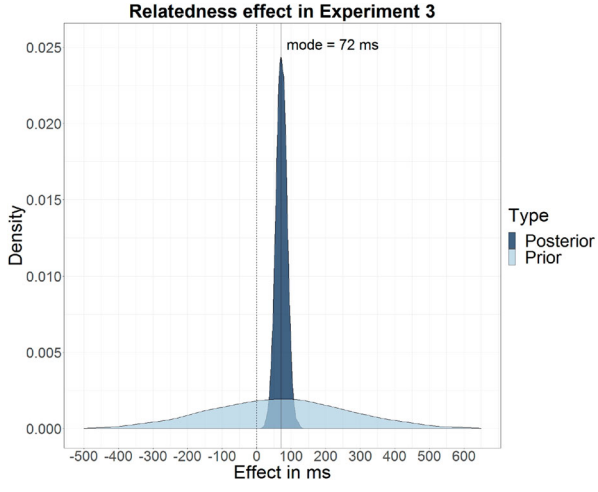

Figure 9. Prior and posterior distributions of Relatedness effect on lexical decision latencies in Experiment 3 drawn from Bayesian linear regression model. Prior distribution was informed by the observed effects of Relatedness in Experiments 1 and 2.

previous results (Barthel et al., 2016; Bögels, Magyari, et al., 2015).

In a quarter of trials, participants had to attend to a switch task midway through preparing their verbal response and make a lexical decision instead of answering the question. The words for lexical decision appeared shortly after participants' gaze moved towards the target picture and were either phonologically related to the verbal response in preparation, i.e. the target picture's name, or not. Phonological relatedness led to longer decision latencies and increased error rates. Importantly, the effect of phonological interference was equally strong in questions giving away the answer in the middle of the question and in questions giving away the answer only at their end, indicating that participants were planning their response phonologically as early as possible and already before the incoming question came to an end.

5. General discussion

Previous research into dialogic turn-taking has shown that next speakers regularly start to plan their turn as early as possible and often in overlap with the incoming turn (Barthel et al., 2016; Bögels, Magyari, et al., 2015). While the timing of speech planning and turn taking is certainly dependent on speakers' communicative intentions and therefore under some amount of strategic control, the early-planning strategy leads to advantages in turn-timing, as next speakers manage to shorten the gaps between turns when they start planning their response in overlap (Barthel et al., 2017; Corps et al., 2018). While a number of studies have shown that next speakers initiate planning in overlap, they did not investigate which steps of response preparation are run through while the current turn is still coming in and potentially interfering with simultaneously running planning processes. Planning in overlap comes at the cost of increased processing load (Barthel & Sauppe, 2019), which might cause next speakers to postpone certain processing stages until the end of the incoming turn in order to avoid high peaks in processing load.

This study investigated which steps of language planning occur in overlap with the incoming turn, and focussed on attesting phonological activation of planned words in the course of the tested experiments. To that end, participants were tested in a switch task combining picture naming and visual lexical decision in a series of three experiments. In Experiment 1, participants had to name pictures as a base task in three quarters of trials and switch to lexical decision instead of naming the picture in one quarter of trials. Effects of associative and phonological relatedness of the pictures lexical decision words on decision performance indicate that participants prepared their verbal response at least until activating the phonological representation of the picture name until they gave a lexical decision. Experiment 2 was designed as a replication of Experiment 1 while participants were presented with background speech while doing the switch task. In that manner, we investigated whether the effects observed in Experiment 1 can be expected to be replicated in a dialogic test situation using the same SOA's. As relatedness effects could not be fully replicated with background speech, and naming latencies were longer as compared to Experiment 1, response planning might have been slowed down by distracting incoming speech. For that reason, Experiment 3 was designed to use a longer SOA of 300 ms.

In Experiment 3, participants were tested in a responsive test situation in which they had to answer questions by naming one of four pictures. The cue to the answer of the questions was located either early during the question or only towards the end of the question. In critical trials, participants again had to switch from giving a verbal response to the question to making a lexical decision instead. The timing of this switch was tied to the beginning of response planning, which we operationalised as eye-gaze towards the target object triggering the presentation of the lexical decision word (Just & Carpenter, 1980). In line with findings in the previous literature, participants were shown to initiate response planning as early as possible, usually in overlap with the incoming turn (Barthel et al., 2017, 2016; Bögels, Magyari, et al., 2015; Corps et al., 2018). Words for lexical decision were presented after phonologically related target pictures or after unrelated pictures. Comparing the effects of relatedness of the (initially intended) verbal responses with the lexical decision words allowed us to draw inferences about the progress of speech planning at the moment the lexical decision is given. Phonological relatedness led to deteriorated lexical decision performance, showing that the word forms of the picture names had been activated by the time the task switched. Critically, this phonological interference effect was shown to be equally strong in the middle of questions and at their end. These results are taken as evidence that next speakers plan their utterance phonologically as early as possible while the incoming turn is still unfolding.

In conclusion, we have shown that language production planning proceeds right through to word form retrieval even during the incoming speech that is being responded to. While naming latencies in the presented experiments are generally longer than average turn-transition times in conversational settings, the attested effects can be taken as informative about the processes of speech planning in conversation, where context, topic familiarity, predictable sequences of actions and the like

speed up turn taking. The presented results support models of the psycholinguistics of dialogue that model response planning as taking place as early as possible (Heldner & Edlund, 2010; Levinson & Torreira, 2015), showing that early planning at least includes the stages of conceptual planning and formulation, even though processing costs in response preparation are higher in overlap with the incoming turn than during the silence between turns (Barthel & Sauppe, 2019). A recent study by Bögels and Levinson (in prep.), measuring tongue movements using ultrasound visualisation, shows that articulatory preparation does not happen as soon as possible but is postponed until articulation becomes immediate. Combining this finding with the present results, the question remains whether early response preparation includes the retrieval and construction of phonetic codes and their translation into motor plans while only the movement of the articulators is postponed or whether phonetic and motor planning stages are postponed and triggered by the incoming turn coming to an end, or possibly by the recognition of turn-final cues (Barthel et al., 2017). At least up to word form retrieval, the time course of production planning in a dialogue situation seems to be very similar to a monological test situation like the picture naming task used in Experiment 1. However, planning seems to be somewhat slower in overlap with the incoming turn as compared to planning in silence due to increased cognitive load when comprehension and response preparation run in parallel.

Ethics

All participants gave prior, written, informed consent to the study. The study was approved by the Ethics Committee of the Faculty of Social Sciences, Radboud University Nijmegen.

Notes

1. However, the effect might have been due to differences in acoustic complexity of the tones vs. the syllables used.
2. Removal of these subjects' data did not change the presented pattern of results, as attested in separate analyses.
3. A guideline for Bayes factor interpretation can be found in Jeffreys (1961), see Kass and Raftery (1995)
4. Removal of this subject's data did not change the presented pattern of results, as attested in separate analyses.
5. Removal of this subject's data did not change the presented pattern of results, as attested in separate analyses.

Acknowledgments

We thank Amie Fairs for useful comments concerning the design of the study, and Antje Meyer for comments on an earlier version of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was financed by the European Research Council Advanced grant nr. 269484 INTERACT awarded to SL and by the Max Planck Institute for Psycholinguistics.

ORCID

Mathias Barthel http://orcid.org/0000-0003-1747-1290

References

Alario, F., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology Section A*, *53*(3), 741–764. doi:10.1080/713755907

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*(328), 1–2. doi:10.3389/fpsyg.2013.00328

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. doi:10.1016/j.jml.2012.11.001

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

Barthel, M., & Sauppe, S. (2019). Speech planning at turn transitions in dialogue is associated with increased processing load. *Cognitive Science*, *43*(7), e12768. doi:10.1111/cogs.12768

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). doi:10.3389/fpsyg.2016.01858

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

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

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*(12881), 1–11. doi:10.1038/srep12881

Bürkner, P. C. (2017). BRMS : An R package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1). doi:10.18637/jss.v080.i01

Clayman, S. (2002). Sequence and solidarity. In *Group cohesion, trust and solidarity* (pp. 229–253). Oxford: Elsevier Science Ltd.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content

predictions to determine what to say but not when to say it. *Cognition*, *175*, 77–95. doi:10.1016/j.cognition.2018.01.015

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.

De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1424 Dutch words in a continuous task. *Behavior Research Methods*, *40*(1), 198–205. doi:10.3758/BRM.40.1.198

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology*. Hove; Hillsdale: L. Erlbaum Associates, Publishers.

Fairs, A., Bögels, S., & Meyer, A. S. (2018). Dual-tasking with simple linguistic tasks: Evidence for serial processing. *Acta Psychologica*, *191*, 131–148. doi:10.1016/j.actpsy.2018.09.006

Fargier, R., & Laganaro, M. (2016). Neurophysiological modulations of non-verbal and verbal dual-tasks interference during word planning. *PloS One*, *11*(12), e0168358. doi:10.1371/journal.pone.0168358

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. doi:10.3758/BF03193857

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

Halekoh, U., & Hojsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568. doi:10.1016/j.wocn.2010.08.002

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. doi:10.1016/j.cognition.2002.06.001

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Jongman, S. R., & Meyer, A. S. (2017). To plan or not to plan: Does planning for production remove facilitation from associative priming? *Acta Psychologica*, *181*, 40–50. doi:10.1016/j.actpsy.2017.10.003

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi:10.1080/01621459.1995.10476572, Retrieved from http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572

Kemper, S., Herman, R. E., & Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging*, *18*(2), 181–192. doi:10.1037/0882-7974.18.2.181

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. doi:10.2307/2533558

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

Klaus, J., Mädebach, A., Oppermann, F., & Jescheniak, J. D. (2017). Planning sentences while doing other things at the same time: Effects of concurrent verbal and visuospatial working memory load. *Quarterly Journal of Experimental Psychology*, *70*(4), 811–831. doi:10.1080/17470218.2016.1167926

La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology*, *81*(4), 511–525. doi:10.1111/j.2044-8295.1990.tb02376.x

Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares means [R package].

Levelt, W. (1989). *Speaking: From intention to articulation*. London: MIT Press.

Levelt, W. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1–3), 1–22. doi:10.1016/0010-0277(92)90038-J

Levelt, W., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–75. doi:10.1017/S0140525X99001776

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122–142.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*(731), 10–26. doi:10.3389/fpsyg.2015.00731

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. doi:10.1016/0010-0277(85)90021-6

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, *53*(2), 146–193. doi:10.1016/j.cogpsych.2006.03.001

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, *66*(3), 180–194. doi:10.1007/s00426-002-0086-5

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. doi:10.1017/S0140525X12001495

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*(2), 291–321. doi:10.1080/01688639508405124

Pomeranz, A., & Heritage, J. (2012). Preference. In T. Stivers & J. Sidnell (Eds.), *The handbook of conversation analysis*. Chichester: Wiley-Blackwell.

Pylkkänen, L., Gonnerman, L., Stringfellow, A., & Marantz, A. (n.d.). *Disambiguating the source of phonological inhibition effects in lexical decision: An MEG study* (p. 27). Submitted.

Rajah, M. N., Languay, R., & Grady, C. L. (2011). Age-related changes in right middle frontal gyrus volume correlate with altered episodic retrieval activity. *Journal of Neuroscience*, *31*(49), 17941–17954. doi:10.1523/JNEUROSCI.1690-11.2011

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., …Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, *15*(11), 1676–1689. doi:10.1093/cercor/bhi044

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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. *Journal of Psycholinguistic Research*, *35*(2), 189–213. doi:10.1007/s10936-005-9011-6

Schriefers, H., Meyer, A. S., & Levelt, W. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102. doi:10.1016/0749-596X(90)90011-N

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

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*, *106*(26), 10587–10592. doi:10.1073/pnas.0903616106

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. doi:10.1126/science.7777863

Wilshire, C., Singh, S., & Tattersall, C. (2016). Serial order in word form retrieval: New insights from the auditory picture-word interference task. *Psychonomic Bulletin & Review*, *23*(1), 299–305. doi:10.3758/s13423-015-0882-8