



# Semantic Context Effects in Picture and Sound Naming: Evidence from Event-related Potentials and Pupillometric Data

Magdalena Gruner<sup>1,2\*</sup>, Andreas Widmann<sup>1\*</sup>, Stefan Wöhner<sup>1</sup>,  
Erich Schröger<sup>1</sup>, and Jörg D. Jescheniak<sup>1</sup>

## Abstract

■ When a picture is repeatedly named in the context of semantically related pictures (homogeneous context), responses are slower than when the picture is repeatedly named in the context of unrelated pictures (heterogeneous context). This semantic interference effect in blocked-cyclic naming plays an important role in devising theories of word production. Wöhner, Mädebach, and Jescheniak [Wöhner, S., Mädebach, A., & Jescheniak, J. D. Naming pictures and sounds: Stimulus type affects semantic context effects. *Journal of Experimental Psychology: Human Perception and Performance*, 47, 716–730, 2021] have shown that the effect is substantially larger when participants name environmental sounds than when they name pictures. We investigated possible reasons for this difference, using EEG and pupillometry. The behavioral data replicated Wöhner and colleagues. ERPs were more positive in the homogeneous compared with the heterogeneous context over central electrode locations between 140–180 msec

and 250–350 msec for picture naming and between 250 and 350 msec for sound naming, presumably reflecting semantic interference during semantic and lexical processing. The later component was of similar size for pictures and sounds. ERPs were more negative in the homogeneous compared with the heterogeneous context over frontal electrode locations between 400 and 600 msec only for sounds. The pupillometric data showed a stronger pupil dilation in the homogeneous compared with the heterogeneous context only for sounds. The amplitudes of the late ERP negativity and pupil dilation predicted naming latencies for sounds in the homogeneous context. The latency of the effects indicates that the difference in semantic interference between picture and sound naming arises at later, presumably postlexical processing stages closer to articulation. We suggest that the processing of the auditory stimuli interferes with phonological response preparation and self-monitoring, leading to enhanced semantic interference. ■

## INTRODUCTION

The use of pictures of objects (e.g., dog) to elicit verbal responses has a long tradition in the chronometric study of speaking, dating back to Cattell in the late 19th century (1886). This approach is highly convenient, as pictures activate specific semantic representations that then serve as input to the speech production process (see Bock, 1996). More recently, environmental sounds (e.g., the bark of a dog) have been used in this context as well, because they also activate semantic representations with high precision (e.g., Kitazawa et al., 2023; Brownsett, Mascelloni, Gowlett, McMahon, & de Zubicaray, 2022; Wöhner et al., 2021; Wöhner, Jescheniak, & Mädebach, 2020; Jeon & Lee, 2009; for studies using environmental sounds for investigating cross-modal semantic integration and memory see, e.g., Mädebach, Kieseler, & Jescheniak, 2018; Mädebach, Wöhner, Kieseler, & Jescheniak, 2017;

Hendrickson, Walenski, Friend, & Love, 2015; Chen & Spence, 2010, 2011; Yuval-Greenberg & Deouell, 2009; Schneider, Engel, & Debener, 2008; Molholm, Ritter, Javitt, & Foxe, 2004).

Many picture-naming studies have focused on semantic context effects; these effects are assumed to be informative with respect to core mechanisms involved in word production (e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992; Schriefers, Meyer, & Levelt, 1990). One prominent task to study such effects is the blocked-cyclic naming (BCN) task in which participants repeatedly name small sets of usually four to five pictures. These sets are either sorted by semantic category (e.g., pictures of a dog, a cow, a pig, and a horse—homogeneous context) or are intermixed (e.g., pictures of a dog, a hammer, a guitar, and a car—heterogeneous context). Each repetition of all items of a set is called a cycle. A well-established finding from BCN studies is that naming responses are slower in the homogeneous context than in the heterogeneous context, typically from the second cycle onward (e.g., Wöhner, Luckow, et al., 2024; Wöhner, Mädebach, Schriefers, &

<sup>1</sup>Leipzig University, <sup>2</sup>University Hospital Carl Gustav Carus, Dresden, Germany

\*Magdalena Gruner and Andreas Widmann contributed equally to this work as co-first authors.

Jescheniak, 2024; Lin, Kuhlen, Melinger, Aristei, & Abdel Rahman, 2022; Wöhner et al., 2021; Abdel Rahman & Melinger, 2007; Belke, Meyer, & Damian, 2005; Damian & Als, 2005; Damian, Vigliocco, & Levelt, 2001). This semantic interference effect is assumed to have its origin at the semantic level and its locus at the lexical level, during the selection of an abstract lexical representation (e.g., Wöhner, Luckow, et al., 2024; Roelofs, 2018; Belke, 2013; cf. Riley, McMahon, & de Zubicaray, 2015). The term origin denotes the processing level at which a change underlying a respective effect occurs, and the term locus refers to the processing level at which the behavioral consequence arises.

Semantic interference is observed not only in the standard BCN task just described (i.e., when pure homogeneous and pure heterogeneous conditions are presented) but also in variants of it when filler trials (with unrelated pictures or an unrelated task) are interspersed or when items from a homogeneous subset and a heterogeneous subset alternate or are randomly intermixed (so that some items appear in a more homogeneous context than others, e.g., Wöhner, Luckow, et al., 2024; Belke, 2013; Damian & Als, 2005). The detrimental effect of having previously named one or more exemplars from a given semantic category is also observed in a related naming task in which pictures are not repeated and no blocking is involved (continuous naming, e.g., Belke, 2013; Howard et al., 2006). These relatively persistent interference effects have led to an important shift in theorizing about word production. Rather than—tacitly—assuming that the network of semantic and lexical representations accessed in word production is largely fixed and invariant, long-term developmental changes notwithstanding (e.g., Levelt et al., 1999; Dell, 1986), more recent models emphasize the adaptive nature of word production (Oppenheim, Dell, & Schwartz, 2010; Howard et al., 2006; Damian & Als, 2005). Specifically, these models assume that the strength of connection in the network is continuously updated with each production episode. This adaptive mechanism is assumed to operate at the interface of semantic and lexical representations. It is modeled in terms of changes in the connection weights between the semantic input (semantic features or holistic concepts) and abstract lexical nodes, which are updated after each single naming episode. Because of these changes, a selected word becomes more accessible for future retrieval, whereas a semantically related competitor word that was coactivated but not selected becomes less accessible for future retrieval.

In a recent BCN study, Wöhner and colleagues (2021) used not only pictures but also sounds. Pictures and sounds were presented intermixed, and participants produced the same target words in response to pictures or sounds (e.g., the word “dog” in response to a picture of that animal or in response to a bark). There were two basic findings. First, sound naming was slower than picture naming, across cycles 736 versus 586 msec (in Experiment 2). Second and more importantly, interference was substantially larger with sounds than with pictures, across

cycles 108 msec versus 34 msec (in Experiment 2); this pattern was replicated with different materials (in Experiment 5) and when only pictures or only sounds were presented in a between-participants design (in Experiment 1). This difference in the magnitude of the interference effect cannot be explained in terms of proportional slowing (i.e., linear scaling of effects). The pattern also persisted when analyses were based on log-transformed latencies or on the relative size of the effect, compensating for differences in performance speed. This difference in the magnitude of semantic interference with pictures and sounds has important implications with respect to the interpretation of semantic context effects in speech production. The data from Wöhner and colleagues support the notion not only that prelexical semantic processes are critically involved in the rise of the interference effect in BCN but also that they may affect its magnitude. This is because in that study, all lexical processing steps (e.g., selection of an abstract lexical node and retrieval of the corresponding phonological form) were identical in picture and sound naming because participants produced the same words to the stimuli. Notably, in both picture and sound naming, the production process starts from an activated semantic representation, which makes processing in these tasks similar to processing involved in spontaneous speech. Moreover, the speaker produces a word (e.g., “dog”) to comment on some relevant event in the external physical world, be it the existence of a perceived entity or a perceived sound. Because of these similarities, the use of sound stimuli can be considered a feasible, useful, and welcome tool for the experimental study of language production. However, how can the difference in the magnitude of semantic interference between pictures and sounds be explained?

It has been argued that sound processing is characterized by coarse-to-fine semantic processing (Murray & Spierer, 2009; Murray, Camen, Spierer, & Clarke, 2008; Murray, Camen, Gonzalez Andino, Bovet, & Clarke, 2006). According to this account, sounds first activate a broader semantic concept like “animal” and only later on a specific concept like “dog.” Although a similar suggestion has been made with respect to pictures (e.g., Mack & Palmeri, 2015; see Clarke, 2015, and Fabre-Thorpe, 2011, for reviews), the temporal progression from coarse-to-fine semantic activation may differ between sounds and pictures. If so, this could mean that—in the context of a naming task—sounds generate a more diffuse activation pattern than pictures at the semantic level, which is then reflected in a more diffuse activation pattern at the lexical level (with multiple abstract lexical nodes being activated to a similar degree). This situation would render selection of the target node at the lexical level more difficult (e.g., Howard et al., 2006; Levelt et al., 1999; Roelofs, 1992), leading to increased semantic interference in sound naming.

The aim of the present study was to test this account. To do so, we replicated Experiment 2 of Wöhner and

colleagues (2021), while recording EEG and pupil diameter data in addition to naming latencies. Because of their high temporal resolution, the EEG data should provide further insight into the processing levels that contribute to the different magnitude of the interference effects in picture and sound naming. As cognitive or attentional demands are known to modulate pupil dilation (e.g., Strauch, Wang, Einhäuser, Van der Stigchel, & Naber, 2022) also in language processing (e.g., Schmidtke, 2018; Papesh & Goldinger, 2012), we hypothesized that the pupillometric measure would also be sensitive to semantic interference in BCN and could thus provide additional information about the issue at hand.

### Electrophysiological Markers of Semantic Context Effects

To the best of our knowledge, there is so far no study investigating ERP correlates of semantic interference in BCN with sounds. However, there are studies investigating ERP correlates of semantic interference in BCN with pictures. These studies found ERP effects in a time window from about 200–500 msec after picture onset (e.g., Python, Fargier, & Laganaro, 2018; Janssen, Hernández-Cabrera, van der Meij, & Barber, 2015; Aristei, Melinger, & Abdel Rahman, 2011; Janssen, Carreiras, & Barber, 2011), sometimes even in an earlier time window (e.g., Lin et al., 2022; Maess, Friederici, Damian, Meyer, & Levelt, 2002). For example, Lin and colleagues (2022) found ERP effects from 140 to 180 msec and from 250 to 350 msec. Against the background of time estimates for subprocesses involved in picture naming derived by Indefrey and Levelt (2004; for an update, see Indefrey, 2011) from a meta-analysis of behavioral and physiological word production studies, these two ERP effects likely reflect semantic processing (conceptual preparation) and subsequent initial lexical processing (lexical selection). As we assumed that the difference in the magnitude of semantic interference in picture and sound naming would be related to these processes, we a priori decided to use the same time windows as Lin and colleagues. However, given that sound naming is slower than picture naming (see Wöhner et al., 2020, for sound naming and Mädebach et al., 2017, for picture naming), it could well be that ERP effects for sounds are delayed relative to ERP effects for pictures.

With respect to the topography and direction of the ERP effects, the majority of BCN studies with pictures either reported an enhanced negativity in the homogeneous context compared with the heterogeneous context at temporal or posterior electrode sites (Lin et al., 2022; Feng, Damian, & Qu, 2021; Aristei et al., 2011; Maess et al., 2002) or an enhanced positivity at frontal or central electrode sites<sup>1</sup> (Wang, Shao, Chen, & Schiller, 2018; Janssen et al., 2011, 2015; Aristei et al., 2011). Possibly, these two effects stem from a dipolar component structure highlighted by average referencing. We therefore defined two ROIs for the present study: a temporoparietal one

and a central one. We predicted ERP effects to be larger for sounds than for pictures, reflecting the difference in the magnitude of the effect at the behavioral level.

### Pupil Size and Semantics

To the best of our knowledge, there are no pupillometric data on semantic interference in BCN with either pictures or sounds. However, there are studies that have used pupillometry in the investigation of other aspects of language processing (for an overview, see Schmidtke, 2018). Pupil dilation was found to be stronger when participants listened to semantically ambiguous sentences than when they listened to unambiguous control sentences (Kadem, Herrmann, Rodd, & Johnsrude, 2020), when participants identified auditory words among a set of written words that included a similar-sounding lexical competitor than when they identified the words among a set of written words that did not include such a competitor (Kuchinsky et al., 2013), when response planning overlapped with incoming turns in a dialog situation (Barthel & Sauppe, 2019), or when participants read aloud low-frequency words than when they read aloud high-frequency words in a delayed naming task (Papesh & Goldinger, 2012). These findings suggest that when people perform more difficult cognitive operations, including semantic and linguistic processes, their pupils dilate (Porter, Troscianko, & Gilchrist, 2007; Kahneman & Beatty, 1966). Because picture naming in BCN takes longer in the homogeneous context than in the heterogeneous context, indicating increased task difficulty, we predicted larger pupil dilation in the homogeneous context than in the heterogeneous context. In addition, because the magnitude of the context effect is larger for sounds than for pictures, we also predicted a larger context effect on pupil dilation for sounds than for pictures.

## METHODS

### Participants

We tested 32 participants (25 female, 7 male;  $M_{\text{age}} = 22.94$  years,  $SD_{\text{age}} = 3.44$  years, range = 18–31 years); most of them were students from Leipzig University. All participants were healthy German native speakers and had normal or corrected-to-normal vision and no known hearing deficits. All but one of the participants were right-handed. They confirmed to have no history of neurological disorders or diseases, not to be pregnant, and not to be under the influence of sedatives or medication affecting the central nervous system. Before the beginning of the experiment, they gave informed written consent. They either received course credits or 8 € per hour. Our study was approved by the ethics committee of Leipzig University (reference number: 20230321\_eb\_188) and followed both the principles of the Declaration of Helsinki and the ethical guidelines of the German Psychological Society (DGPs). Our study builds on previous evidence collected

in studies with young healthy adults (predominantly university students). We recruited participants from a similar population to ensure comparability with these previous studies. We report all data exclusions, manipulations, and measures implemented. We preregistered our study: [https://aspredicted.org/G8D\\_2MY](https://aspredicted.org/G8D_2MY). Note that, in deviation from preregistration, we did not perform exploratory factor analysis and cluster-based permutation tests, but time window mean analyses. This approach was chosen to improve comparability of our results to the studies we directly refer to (Lin et al., 2022; Janssen et al., 2015; Aristei et al., 2011). Data were collected between April and June 2023.

### Power Considerations

We determined the participant sample size based on two criteria: first, on the sample sizes in previous studies investigating ERP correlates of semantic interference in BCN (typically 20–32 participants) and, second, on the number of participants needed for the counterbalancing of the experimental conditions. For the behavioral data, we also estimated the power of our experiment to obtain a similar interaction of Context and Stimulus Type as Wöhner and colleagues (2021; Experiment 2) by means of simulation. This calculation was based on the respective data from Wöhner and colleagues (including all cycles) and done with the Superpower package (Lakens & Caldwell, 2021, 2022). Simulated cell means were 600, 570, 790, and 685 msec for pictures presented in the homogeneous context, pictures presented in the heterogeneous context, sounds presented in the homogeneous context, and sounds presented in the heterogeneous context, respectively. Corresponding standard deviations were 60, 50, 125, and 100 msec for participants and 30, 30, 130, and 85 msec for items. Correlations between repeated measures varied between .7 and .9 for participants and for items. The simulation revealed that the design of our experiment with 32 participants and 16 items has a power of >99% in the participant analysis and 98% in the item analysis to observe the interaction of Semantic Context and Stimulus Type in an ANOVA (for a two-sided test with  $\alpha = .05$ ). Assuming that linear mixed-model (LMM) analyses are at least as sensitive as ANOVAs, the power of our design should be sufficiently high for these analyses as well.

Note that our study included twice as many cycles as Wöhner and colleagues (eight instead of four), and Cycle 1 (which typically does not contribute to interference or only to a lesser extent and also did so in our study; see below) was excluded from the analysis. For these reasons, the true power of our experiment for finding the respective interaction in the behavioral data is likely even higher.

### Materials

We used the materials from Wöhner and colleagues (2021; Experiment 2). The pictures were color photographs of 16

common objects, four each from four semantic categories (tools, vehicles, mammals, and musical instruments). They filled an imaginary square of 250 pixels  $\times$  250 pixels (corresponding to a visual angle of  $7.0^\circ \times 7.0^\circ$  at about 70-cm viewing distance) and had their background removed. For each picture, there was a corresponding sound (e.g., a bark for the picture of a dog). The duration of the sounds ranged from 796 to 1448 msec ( $M = 1143$  msec,  $SD = 218$  msec).

Stimuli were arranged in a  $4 \times 4$  matrix such that each row contained a subset of items used for creating a homogeneous block and each column contained a subset of items used for creating a heterogeneous block (see Appendix). Thus, there were four homogeneous blocks and four heterogeneous blocks. To collect a sufficient number of observations per participant and experimental condition for the EEG analyses, we created eight cycles (compared with four in Wöhner et al., 2021, Experiment 2; the interference effect in picture naming is known to be stable across cycles from Cycle 2 onward; see, e.g., Wöhner, Luckow, et al., 2024, which used up to 12 cycles). This resulted in 128 observations per context condition (homogeneous vs. heterogeneous) and stimulus type (picture vs. sound) per participant. In the eight trials of a cycle, each item of the respective subset was presented once as picture and once as sound. In the first four trials, two of the four items were presented as pictures and the other two items as sounds; in the last four trials, the stimulus type for the items was changed. This was counterbalanced by creating two parallel versions of experimental lists. As a consequence, no more than four stimuli of the same type (picture vs. sound) could occur in adjacent trials. We created 16 such pairs of pseudorandomized lists with mix (van Casteren & Davis, 2006) with the constraint that no item would be repeated in adjacent trials. Blocks generated from the subsets of items in row<sub>*n*</sub> and column<sub>*n*</sub> were combined into four pairs and were always presented in direct succession. The sequence of these four pairs was controlled across participants using a sequentially balanced Latin square design. For each participant, homogeneous and heterogeneous blocks alternated. Half of the participants started with a homogeneous block and the other half with a heterogeneous block.

### Apparatus

Pictures were presented in the center of a 24-in. monitor (ViewPixx/EEG, VPixx Technologies). Screen resolution was 1920  $\times$  1080 pixel, background color was gray (RGB 73 73 73, 15.8 cd/m<sup>2</sup>), and viewing distance was about 70 cm. Sounds were presented with Sennheiser HD25 headphones ( $M = 70.5$  dB SPL, range = 63–75.5 dB SPL). The experiment was controlled by GNU Octave and the Psychophysics Toolbox (Kleiner et al., 2007; Brainard, 1997; Pelli, 1997) operated under Linux. Participants' naming responses were registered with a Sennheiser K6/ME 64 microphone and digitally recorded for a time window of



3 sec starting from stimulus onset. The EEG was recorded at a sampling rate of 500 Hz with 32 active electrodes using a BrainAmp EEG amplifier (Brainproducts). Electrodes were placed according to a subset of the 10–10 system and left and right mastoids and online referenced to an electrode on the tip of the nose. Eye movements were recorded by placing two electrodes on the outer canthi and one below the left eye. Pupil diameter was recorded from both eyes with an infrared EyeLink Portable Duo eye-tracker (SR Research Ltd.). It was configured to operate in remote mode, capturing the pupil diameter at a sampling rate of 500 Hz.

## Design

Context (homogeneous vs. heterogeneous), stimulus type (picture vs. sound), and cycle (1–8) were tested within participants and within items.

## Procedure

The experiment was conducted in the BioCog research laboratory at Leipzig University. Participants were tested individually and were seated in a comfortable office chair in an electrically shielded, double-walled sound booth (Industrial Acoustics Company). First, participants received written instructions. Then, they were familiarized with the stimuli. A picture and the corresponding sound were presented together (picture centered on the screen for 3 sec), and 1 sec after onset of the two stimuli, their name appeared in the lower part of the screen. Participants were asked to read the name aloud and to use only that name in the experiment. Next, there was a practice block. Each picture–sound pair was presented once more, but this time without the name. Participants were asked to produce the corresponding name quickly and accurately. Wrong responses were corrected by the experimenter when the practice block was completed. Then eight experimental blocks (with 64 trials each) started. In each experimental trial, either a picture or a sound was presented. There were short breaks between experimental blocks.

Experimental trials started with a black fixation cross (400 msec), followed by a blank screen (400 msec) and either the target picture (1500 msec) or the target sound (796–1448 msec). Trials lasted 4 sec.

At the end of the experiment, there was an extra block in which participants named the stimuli with a delay (see Mädebach, Widmann, Posch, Schröger, & Jescheniak, 2022). We included this block to facilitate the identification of speech artifacts during EEG preprocessing. In this block, each item was presented twice as picture and twice as sound in intermixed order, resulting in 64 delayed naming trials. Stimuli were presented as in the experimental trials, but there was an additional response cue (black square, sized to frame the picture stimuli) that appeared on the screen 1200 msec after stimulus onset. Participants were instructed to prepare the naming response swiftly

when the target stimulus was presented (as they had done before), but to initiate articulation only when the response cue appeared. We added 12 catch trials, in which the response cue appeared with a short delay (of 600 msec) to motivate participants to prepare the naming response as soon as the stimulus was presented, rather than simply waiting for the response cue.

## Data Preprocessing

### *Behavioral Data*

Naming latencies and naming correctness were determined offline by means of visual and auditory inspection using CheckFiles (Protopapas, 2007). Cases in which participants produced no response, an unexpected response, or a disfluent response (1.92% of all observations) were coded as participant errors and excluded from the analyses of naming latencies, EEG, and pupil dilation. Trials were also excluded if the naming latency was shorter than 300 msec (0.34% of all observations). Because participant errors were so rare, we do not report a statistical analysis of them in this article. However, it is important to note that there was no evidence of a speed-accuracy trade-off. For details on the error data, see <https://osf.io/9hjtq/>.

### *EEG Data*

EEG data were preprocessed using EEGLAB (Delorme & Makeig, 2004). First, we applied a 75-Hz low-pass filter (163 point-filter, transition band width = 10 Hz; Widmann, Schröger, & Maess, 2015) and a 0.1-Hz high-pass filter (8025 point-filter, transition band width = 0.2 Hz). The data were then divided into epochs of 1200 msec time-locked to the target onset, ranging from 200 msec before stimulus onset to 1000 msec after stimulus onset. Trials with peak-to-peak amplitudes >1000  $\mu$ V were excluded to eliminate trials with nonstereotypical artifacts, because of, for example, swallowing, but retain stereotypical artifacts, because of, for example, speech or eye blinks (to be later removed from the data by means of independent component analysis [ICA]). Channels were excluded if their robust  $z$  score of the robust standard deviation was higher than 3. For the removal of speech artifacts and other artifacts, we performed an ICA, using the AMICA algorithm (Delorme, Palmer, Onton, Oostenveld, & Makeig, 2012). For the ICA, the data were low-pass filtered with the same low-pass filter as described above, but high-pass filtered with a 1-Hz high-pass filter (1605 point-filter, transition band width = 1 Hz). The same bad channels and epochs as above were excluded, and the data were divided into epochs of 1.2 sec (–200 to 1000 msec relative to target onset). Independent components were classified in EEGLAB. Eye movement components (presaccadic spike potentials and horizontal and vertical movement of corneo-retinal dipoles), blinks, muscle artifact components, speech artifact components, and other noise

components were determined manually after classification with ICLABEL EEGLAB plugin (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019). The manual classification considered the topography of components (focal vs. widespread), peak in the alpha frequency range (8–12 Hz), and presence of stimulus evoked activity. Speech artifact components were identified as components that occurred only in experimental trials but not in the delayed naming trials. Components were selected manually rather than automatically to minimize the exclusion of components that reflect neural activity. The selected artifact IC activity was then subtracted from the data. We removed an average of 18.13 components per participant ( $Mdn = 17$ ,  $min = 10$ ,  $max = 24$ ). Excluded bad channels were interpolated employing spherical spline interpolation. Data were baseline-corrected using the 200-msec prestimulus interval. Epochs with peak-to-peak amplitudes higher than 150  $\mu V$  were excluded. We calculated individual average ERPs per participant, context, and stimulus type. All channels were rereferenced to “infinity” using the reference electrode standardization technique developed by Yao (2001). Finally, the residue iteration decomposition (RIDE MATLAB toolbox; Ouyang, Sommer, & Zhou, 2015) was applied. This method allows to decompose the ERP into component clusters with different latency variability. We used residue iteration decomposition to separate the signal into the stimulus-locked S-component and the response-locked R-component. The R-component was subtracted from the ERP to remove any remaining speech artifacts, as well as the exclusively response-related cognitive activity from the brain signal. Statistical analyses were thus performed on the extracted S-component.

For the ERP analyses, we a priori decided to analyze the time windows and ROIs (as close as possible) for which Lin and colleagues (2022) reported semantic context effects in the ERP, namely, 140–180 msec at a temporoparietal ROI (P7, P8, M1, M2) and 250–350 msec at both mastoids (M1, M2). In addition, we analyzed the ERP data in the same time windows at a central ROI (Fz, FC1, FC2, Cz, CP1, CP2, Pz), presumably reflecting the polarity-inverted (positive) aspects of the temporoparietally negative effects as described by Lin and colleagues (2022). After visual inspection of the ERP data, we decided post hoc to analyze two more time windows closer to articulation: 400–500 msec and 500–600 msec at a frontal ROI (F3, Fz, F4, FC1, FC2). These additional analyses were not theory driven, but rather exploratory.

### Pupil Data

Pupil diameter measurements were transformed from Eyelink digital count units to millimeters as proposed by Steinhauer, Bradley, Siegle, Roecklein, and Dix (2022). Blink and eye saccade information were obtained from the eye tracker. During postprocessing, a custom function was applied to detect partial blinks from the smoothed velocity time series by identifying pupil diameter changes

exceeding 20 mm/sec, including a 50-msec preblink and a 100-msec postblink interval, as suggested by Merritt, Keegan, and Mercer (1994). We averaged the data measured from both eyes using the dynamic offset algorithm (Kret & Sjak-Shie, 2019). Blinks and other intervals with signal loss exceeding 1 sec were removed from the data. Shorter periods were interpolated using the MATLAB 1-D interpolation function with shape-preserving piecewise cubic interpolation. The data were then segmented into 4-sec epochs, including a 200-msec prestimulus baseline period. To correct for baseline differences, the epochs were baseline-corrected by subtracting the mean amplitude from the baseline period (–200 to 0 msec). Individual mean pupil dilation responses (PDRs) were calculated for each participant and condition. For statistical analysis, mean PDRs were investigated from the time window 1200–2000 msec after target onset. Given the lack of an established standard in the literature, we opted for a peak-centered time window and consider our analyses exploratory.

### Statistical Analyses

All analyses were performed in R (Version 4.3.1; R Core Team, 2023). We performed LMM analyses on each of the three measures (ERP data, PDR data, and behavioral data) using the R packages *lme4* (Version 1.1–35; Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Version 3.1–3; Kuznetsova, Brockhoff, & Christensen, 2020). Trials were treated as primary unit of investigation (Level 1) nested within participants and items (Level 2). Only trials including valid estimates for ERP as well as PDR amplitudes per trial from Cycles 2–8 (see below for discussion) were included in the LMM analyses. Naming latencies were log-transformed. Although there is no generally optimal transformation of RTs for LMMs and the analysis of untransformed data may be preferable for mental chronometry applications (for discussions, see, Lo & Andrews, 2015; Baayen & Milin, 2010; Wagenmakers & Brown, 2007), here, we log-transformed the naming latencies to avoid misinterpreting a proportional slowing of sound naming as an interaction effect (as similarly done by Wöhner et al., 2021). Participants and items were treated as random variables. All models included fixed effects from dummy-coded variables for Context (homogeneous = 0; heterogeneous = 1), Stimulus Type (picture = 0; sound = 1), and the interaction of them. Simple effects are reported. Adding cycle as a covariate did not improve any of the model fits; we therefore do not report them here. As random effects, we included by-participant and by-item intercepts and slopes for both context and stimulus type effects. Including the Stimulus Type  $\times$  Context interaction random effect for participants or items resulted in singular or nonconverging models. All models were fit using the maximum likelihood method (i.e., REML = FALSE). We report regression coefficients as fixed effects estimates,  $t$  values, 95% confidence intervals (CIs), and

*p* values, derived from the *lmerTest* package. Degrees of freedom were estimated via Satterthwaite's degrees of freedom method. For the full model output for all models, see <https://osf.io/9hjtq/>. Model comparisons were performed based on likelihood ratio tests. We additionally report the Akaike's information criterion (AIC), the Bayesian information criterion (BIC), as well as the marginal  $R_m^2$  (describing the proportion of variance explained by the fixed effects) and the conditional  $R_c^2$  (describing the proportion of variance explained by both the fixed and random effects) from the R package *MuMin* (Version 1.47.5; Bartón, 2023) in Table 1 (ERPs per time window

and PDRs) and Table 2 (behavioral data), respectively. The research question was to identify potential correlates of the enhanced effect of the homogeneous versus heterogeneous context for sound naming compared with picture naming. Therefore, we systematically compared the models including only the two main effects of Context and Modality with the models including the main effects and their interaction. To explore whether ERP and PDR amplitudes could serve as predictors of naming latencies, we compared models including versus excluding the ERP and PDR time window mean amplitudes as continuous covariate in the Stimulus Type  $\times$  Context interaction. All

**Table 1.** AIC, BIC,  $R_m^2$ , and  $R_c^2$  Model Fit Estimates, and Likelihood Ratio Tests for the ERP and PDR Data Model Comparisons Excluding versus Including the Stimulus Type  $\times$  Context Interaction Effect

	AIC	BIC	$R_m^2$	$R_c^2$	$\chi^2$	df	<i>p</i>
ERP 140–180 msec							
Central ROI							
Stimulus Type + Context	82186	82306	0.046	0.213			
<b>Stimulus Type <math>\times</math> Context</b>	<b>82182</b>	<b>82310</b>	<b>0.046</b>	<b>0.213</b>	<b>5.53</b>	<b>1</b>	<b>.019</b>
Temporoparietal ROI							
Stimulus Type + Context	68887	69007	0.019	0.307			
<b>Stimulus Type <math>\times</math> Context</b>	<b>68883</b>	<b>69011</b>	<b>0.019</b>	<b>0.307</b>	<b>5.52</b>	<b>1</b>	<b>.019</b>
ERP 250–350 msec							
Central ROI							
<b>Stimulus Type + Context</b>	<b>83400</b>	<b>83520</b>	<b>0.059</b>	<b>0.202</b>			
Stimulus Type $\times$ Context	83401	83529	0.059	0.202	1.02	1	0.313
Mastoidal ROI							
<b>Stimulus Type + Context</b>	<b>81975</b>	<b>82095</b>	<b>0.015</b>	<b>0.141</b>			
Stimulus Type $\times$ Context	81977	82104	0.015	0.141	0.6	1	0.441
ERP 400–500 msec, frontal ROI							
Stimulus Type + Context	88554	88674	0.004	0.059			
<b>Stimulus Type <math>\times</math> Context</b>	<b>88542</b>	<b>88670</b>	<b>0.005</b>	<b>0.06</b>	<b>13.53</b>	<b>1</b>	<b>&lt;.001</b>
ERP 500–600 msec, frontal ROI							
Stimulus Type + Context	90061	90180	0.017	0.062			
<b>Stimulus Type <math>\times</math> Context</b>	<b>90024</b>	<b>90152</b>	<b>0.019</b>	<b>0.065</b>	<b>38.09</b>	<b>1</b>	<b>&lt;.001</b>
PDR 1200–2000 msec							
Stimulus Type + Context	7623	7743	0.194	0.316			
<b>Stimulus Type <math>\times</math> Context</b>	<b>7579</b>	<b>7706</b>	<b>0.193</b>	<b>0.318</b>	<b>45.92</b>	<b>1</b>	<b>&lt;.001</b>

The preferred model is shown in **bold**.

**Table 2.** AIC, BIC,  $R_m^2$ ,  $R_c^2$  Model Fit Estimates, and Likelihood Ratio Tests for the Naming Latency Data Model Comparisons Excluding versus Including the Stimulus Type  $\times$  Context Interaction Effect and ERP and PDR Interaction Effect, Respectively

	AIC	BIC	$R_m^2$	$R_c^2$	$\chi^2$	df	p
Naming latencies							
Stimulus Type + Context	-3169	-3049	0.208	0.469			
<b>Stimulus Type <math>\times</math> Context</b>	<b>-3402</b>	<b>-3275</b>	<b>0.217</b>	<b>0.479</b>	<b>235.38</b>	<b>1</b>	<b>&lt;.001</b>
Naming latencies and ERP 400–500 msec, frontal ROI							
Stimulus Type $\times$ Context	-3402	-3275	0.217	0.479			
<b>Stimulus Type <math>\times</math> Context <math>\times</math> ERP<sub>400–500 msec</sub></b>	<b>-3543</b>	<b>-3386</b>	<b>0.224</b>	<b>0.483</b>	<b>148.95</b>	<b>4</b>	<b>&lt;.001</b>
Naming latencies and ERP 500–600 msec, frontal ROI							
Stimulus Type $\times$ Context	-3402	-3275	0.217	0.479			
<b>Stimulus Type <math>\times</math> Context <math>\times</math> ERP<sub>500–600 msec</sub></b>	<b>-3548</b>	<b>-3391</b>	<b>0.225</b>	<b>0.481</b>	<b>153.92</b>	<b>4</b>	<b>&lt;.001</b>
Naming latencies and PDR							
Stimulus Type $\times$ Context	-3402	-3275	0.217	0.479			
<b>Stimulus Type <math>\times</math> Context <math>\times</math> PDR</b>	<b>-3751</b>	<b>-3594</b>	<b>0.235</b>	<b>0.488</b>	<b>356.67</b>	<b>4</b>	<b>&lt;.001</b>

The preferred model is shown in **bold**.

continuous covariates were group mean centered. We provide the estimated means of linear trends between ERP amplitudes and naming latencies and between pupil dilation and naming latencies in each stimulus type and context using the R function *emtrends* from the *emmeans* package (Version 1.10.1; Lenth, 2023) in Table 2.

The left column of Figure 3 shows mean naming latencies per stimulus type, context, and cycle. The right column of Figure 3 shows context effects per stimulus type and cycle. Visual inspection of the data reveals that there was a relatively small effect of context in the first cycle (for pictures: 18 msec, 95% CI [-2, 39]; for sounds: 40 msec, 95% CI [3, 76]) and a much larger one from Cycle 2 onward (for pictures: 45 msec, 95% CI [40, 50]; for sounds: 145 msec, 95% CI [134, 156]). This pattern is in line with results from previous BCN experiments with pictures, which typically reported no effect of context or even facilitation in Cycle 1 and stable interference from Cycle 2 onward (see Belke, 2013, for an overview). Specifically, when the first cycle was excluded from the analysis, there was usually no significant interaction of Context and Cycle (see Belke & Stielow, 2013). The experimental design of our study slightly deviates from standard BCN with respect to the definition of a cycle. In standard BCN, one presentation of each item of the set defines a cycle. Because our study introduced stimulus type (picture vs. sound) as an additional variable, two presentations of each item (once as picture and once as sound) define a cycle. Because of this special feature of our design, the data from Cycle 1 likely reflect a mixture of no effect or facilitation on the

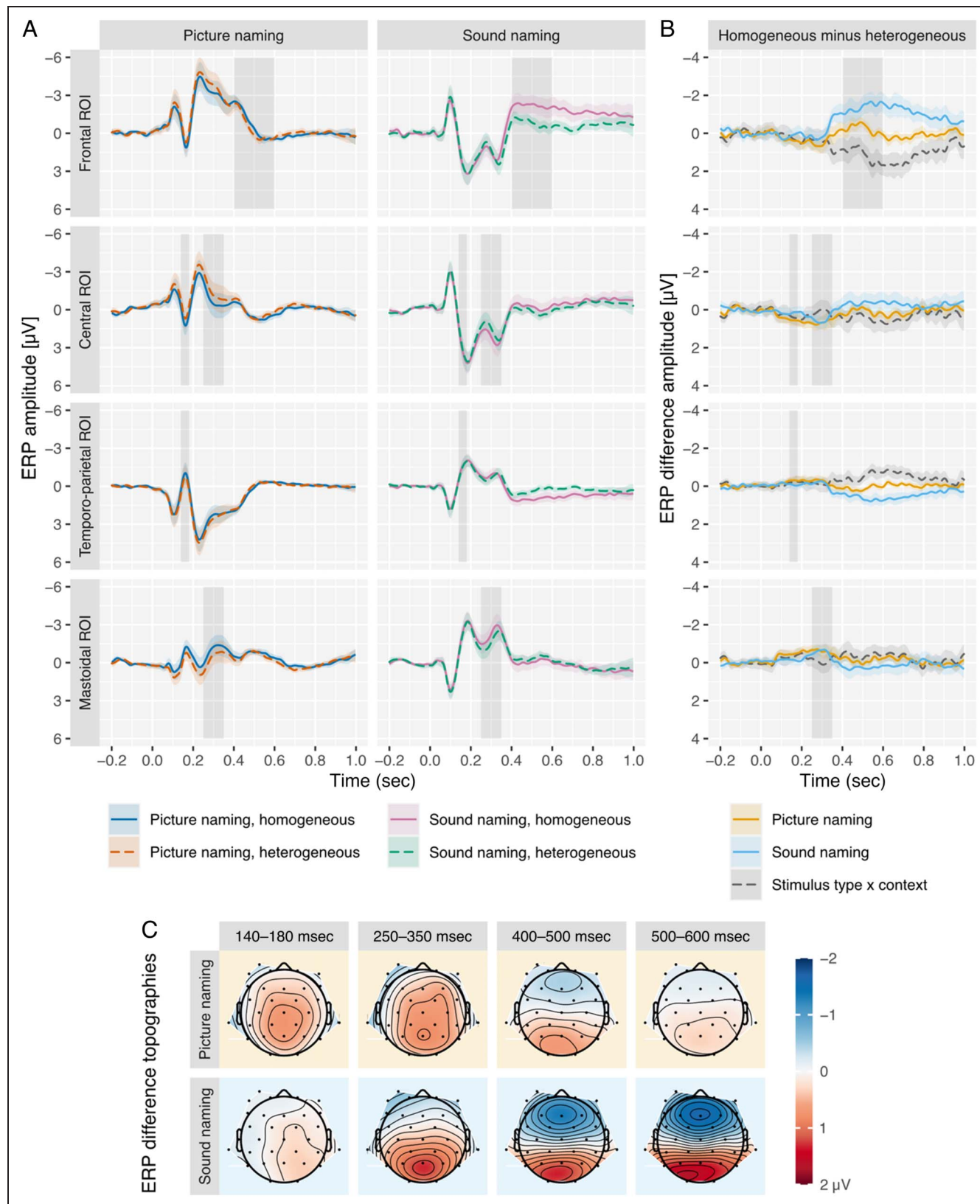
one hand and interference on the other. Figure 3 also shows that there was substantial repetition priming from Cycle 1 to Cycle 2, after which naming latencies remained relatively stable. For these two reasons and in line with previous studies that either excluded the data from Cycle 1 from the analysis (e.g., Wang et al., 2018; Abdel Rahman & Melinger, 2011; Belke, Brysbaert, Meyer, & Ghyselinck, 2005) or analyzed them separately (Lin et al., 2022), we excluded the Cycle 1 data from all analyses reported below and did not consider cycle as an experimental variable. This also gave us sufficient observations for the analysis of the EEG and pupil data per participant and per experimental condition, resulting from the crossing of the two experimental variables we were interested in, namely, stimulus type and context.

## RESULTS

### ERP Data

Figure 1 displays the results of the ERP analyses, including the topographical distribution of the differences between homogeneous minus heterogeneous contexts for picture and sound naming. First, we describe the results from the two time windows we defined based on Lin and colleagues (2022), the early time window at central and temporo-parietal regions and the later time window at central and mastoidal regions. Then, we describe the results from the two additional time windows closer to articulation at frontal regions.





**Figure 1.** (A) ERP waveforms at frontal (F3, Fz, F4, FC1, FC2), central (Fz, FC1, FC2, Cz, CP1, CP2, Pz), temporoparietal (P7, P8, M1, M2), and mastoidal ROIs (M1, M2) for both contexts for picture naming (left) and sound naming (right). (B) Homogeneous minus heterogeneous context difference wave for picture naming and sound naming and the difference of difference waves (reflecting the Stimulus Type  $\times$  Context interaction effect). Shaded areas reflect 95% CIs. Dark gray rectangles indicate the analysis time windows per ROI. (C) Homogeneous minus heterogeneous context difference topographies for picture naming and sound naming in the four analysis time windows.

### ERP Time Window 140–180 msec

In the 140- to 180-msec time window, we observed an effect of context for pictures but not for sounds in both ROIs. The model including the Stimulus Type  $\times$  Context interaction effect fitted the data better than the model including only the main effects in both ROIs (model fit estimates are reported in Table 1). In the central ROI, the ERP amplitudes for pictures were significantly more positive in the homogeneous context than in the heterogeneous context (0.74 vs. 0.18  $\mu\text{V}$ ; Context effect for pictures 0.56 [0.25, 0.88]  $\mu\text{V}$ ,  $t(40.51) = 3.48$ ,  $p = .001$ ), whereas, for sounds, they were similar between the homogeneous context and the heterogeneous context (3.06 vs. 2.93  $\mu\text{V}$ ; Context effect for sounds 0.13 [−0.19, 0.45]  $\mu\text{V}$ ,  $t(41.53) = 0.8$ ,  $p = .431$ ; Stimulus Type  $\times$  Context interaction effect 0.43 [0.07, 0.8]  $\mu\text{V}$ ,  $t(13147.56) = 2.35$ ,  $p = .019$ ). In the temporoparietal ROI, the ERP amplitudes for pictures were significantly more negative in the homogeneous context than in the heterogeneous context (−0.55 vs. −0.24  $\mu\text{V}$ ; Context effect for pictures −0.31 [−0.55, −0.08]  $\mu\text{V}$ ,  $t(35.71) = -2.59$ ,  $p = .014$ ), whereas, for sounds, they were again similar between the homogeneous context and the heterogeneous context (−1.46 vs. −1.41  $\mu\text{V}$ ; Context effect for sounds −0.05 [−0.29, 0.19]  $\mu\text{V}$ ,  $t(36.38) = -0.427$ ,  $p = .672$ ; Stimulus Type  $\times$  Context interaction effect −0.26 [−0.48, −0.04]  $\mu\text{V}$ ,  $t(13121.4) = -2.35$ ,  $p = .019$ ).

### ERP Time Window 250–350 msec

In the 250- to 350-msec time window, we observed an effect of context of similar amplitude between both pictures and sounds in both ROIs. The model including only the Stimulus Type and Context main effects fitted the data better than the model including also the interaction effect in both ROIs. In the central ROI, the ERP amplitudes were significantly more negative for pictures than for sounds (−1.08 vs. 1.87  $\mu\text{V}$ , Stimulus Type effect −2.96 [−3.97, −1.94]  $\mu\text{V}$ ,  $t(39.15) = -5.71$ ,  $p < .001$ ) and significantly more positive in the homogeneous context than in the heterogeneous context (0.7 vs. 0.09  $\mu\text{V}$ , Context effect 0.61 [0.35, 0.86]  $\mu\text{V}$ ,  $t(22.79) = 4.73$ ,  $p < .001$ ). In the mastoidal ROI, the ERP amplitudes were significantly less negative for pictures than for sounds (−0.7 vs. −1.96  $\mu\text{V}$ , Stimulus Type effect 1.26 [0.48, 2.04]  $\mu\text{V}$ ,  $t(39) = 3.18$ ,  $p = .003$ ) and significantly more negative in the homogeneous context than in the heterogeneous context (−1.62 vs. −1.03  $\mu\text{V}$ , Context effect −0.59 [−0.8, −0.38]  $\mu\text{V}$ ,  $t(17.55) = -5.47$ ,  $p < .001$ ).

### ERP Time Window 400–500 msec

In the 400- to 500-msec time window in the frontal ROI, we observed an effect of context for sounds but not for pictures. The model including the Stimulus Type  $\times$  Context interaction effect fitted the data better than the model including only the main effects. The ERP amplitudes for

pictures were similar between the homogeneous context and the heterogeneous context (−1.59 vs. −1.23  $\mu\text{V}$ ; Context effect for pictures −0.36 [−0.8, 0.08]  $\mu\text{V}$ ,  $t(45.12) = -1.61$ ,  $p = .113$ ), whereas, for sounds, they were significantly more negative in the homogeneous context than in the heterogeneous context (−2.31 vs. −1.08  $\mu\text{V}$ , Context effect for sounds −1.23 [−1.67, −0.79]  $\mu\text{V}$ ,  $t(46.24) = -5.46$ ,  $p < .001$ ; Stimulus Type  $\times$  Context interaction effect 0.87 [0.41, 1.33]  $\mu\text{V}$ ,  $t(13151.9) = 3.68$ ,  $p < .001$ ).

### ERP Time Window 500–600 msec

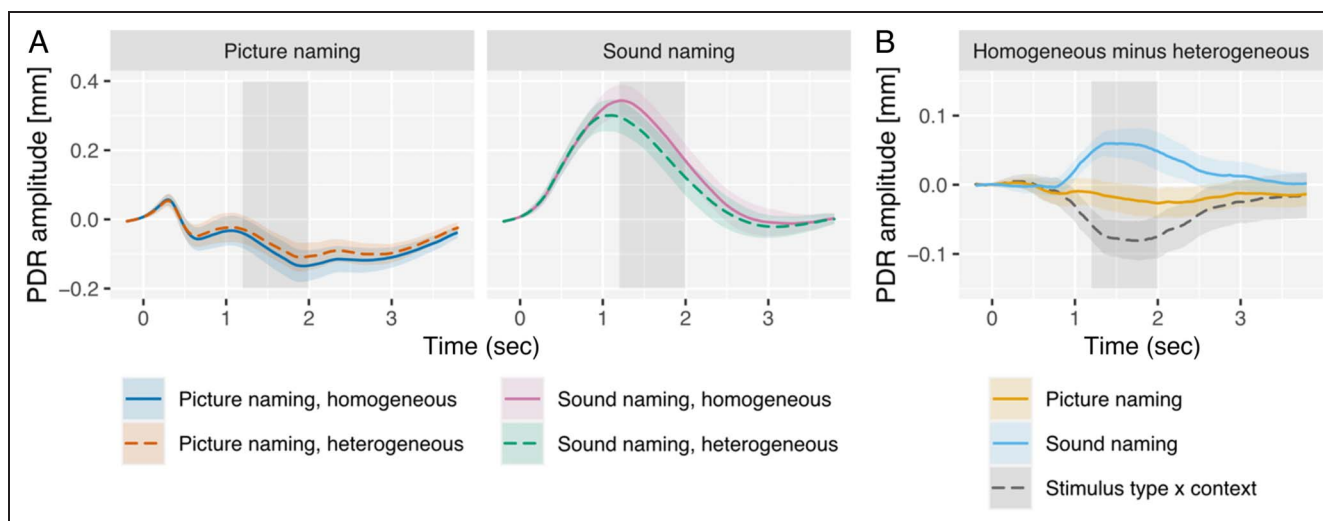
In the 500- to 600-msec time window in the frontal ROI, we observed an effect of Context for sounds but not for pictures. The model including the Stimulus Type  $\times$  Context interaction effect fitted the data better than the model including only the main effects. The ERP amplitudes for pictures were similar between the homogeneous context and the heterogeneous context (0.34 vs. 0.43  $\mu\text{V}$ ; Context effect for pictures −0.09 [−0.58, 0.4]  $\mu\text{V}$ ,  $t(34.81) = -0.36$ ,  $p = .724$ ), whereas, for sounds, they were significantly more negative in the homogeneous context than in the heterogeneous context (−2.16 vs. −0.53  $\mu\text{V}$ ; Context effect for sounds −1.63 [−2.12, −1.14]  $\mu\text{V}$ ,  $t(35.54) = -6.49$ ,  $p < .001$ ; Stimulus Type  $\times$  Context interaction effect 1.54 [1.05, 2.03]  $\mu\text{V}$ ,  $t(13155.1) = 6.18$ ,  $p < .001$ ).

### PDR Data

The results of the PDR data analysis are displayed in Figure 2. In the 1200- to 2000-msec time window, we observed pupil constriction relative to baseline in response to pictures and pupil dilation relative to baseline in response to sounds and an effect of Context for sounds but not for pictures. The model including the Stimulus Type  $\times$  Context interaction effect fitted the data better than the model including only the main effects. The PDR amplitudes for pictures were similar between the homogeneous context and the heterogeneous context (−0.1 vs. −0.08 mm; Context effect for pictures −0.02 [−0.04, 0.001] mm,  $t(45.1) = -1.89$ ,  $p = .065$ ), whereas, for sounds, they were significantly higher in the homogeneous context than in the heterogeneous context (0.28 vs. 0.22 mm, Context effect for sounds 0.06 [0.04, 0.08] mm,  $t(46.4) = 5.59$ ,  $p < .001$ ; Stimulus Type  $\times$  Context interaction effect −0.07 [−0.1, −0.05] mm,  $t(13120) = 6.78$ ,  $p < .001$ ).

### Behavioral Data

The results of the naming latency analysis are displayed in Figure 3. Naming latencies were slower in the homogeneous context than in the heterogeneous context for both sounds and pictures, and also slower for sounds than for pictures in both contexts. The effect of context was larger for sounds than for pictures. The model including the Stimulus Type  $\times$  Context interaction effect fitted the data

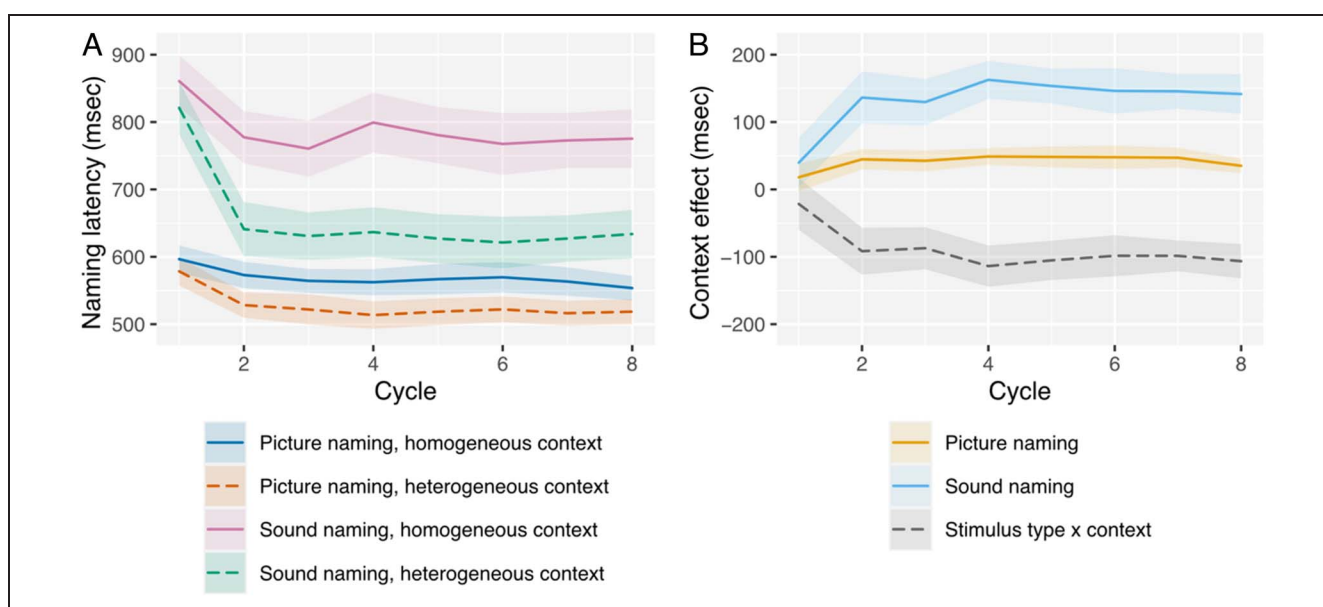


**Figure 2.** (A) PDRs (pupil diameter change relative to baseline) for both contexts for picture naming (left) and sound naming (right). (B) Homogeneous minus heterogeneous context PDR difference for picture naming and sound naming and the difference of PDR differences (reflecting the Stimulus Type  $\times$  Context interaction effect). Shaded areas reflect 95% CIs. Dark gray rectangles indicate the analysis time window.

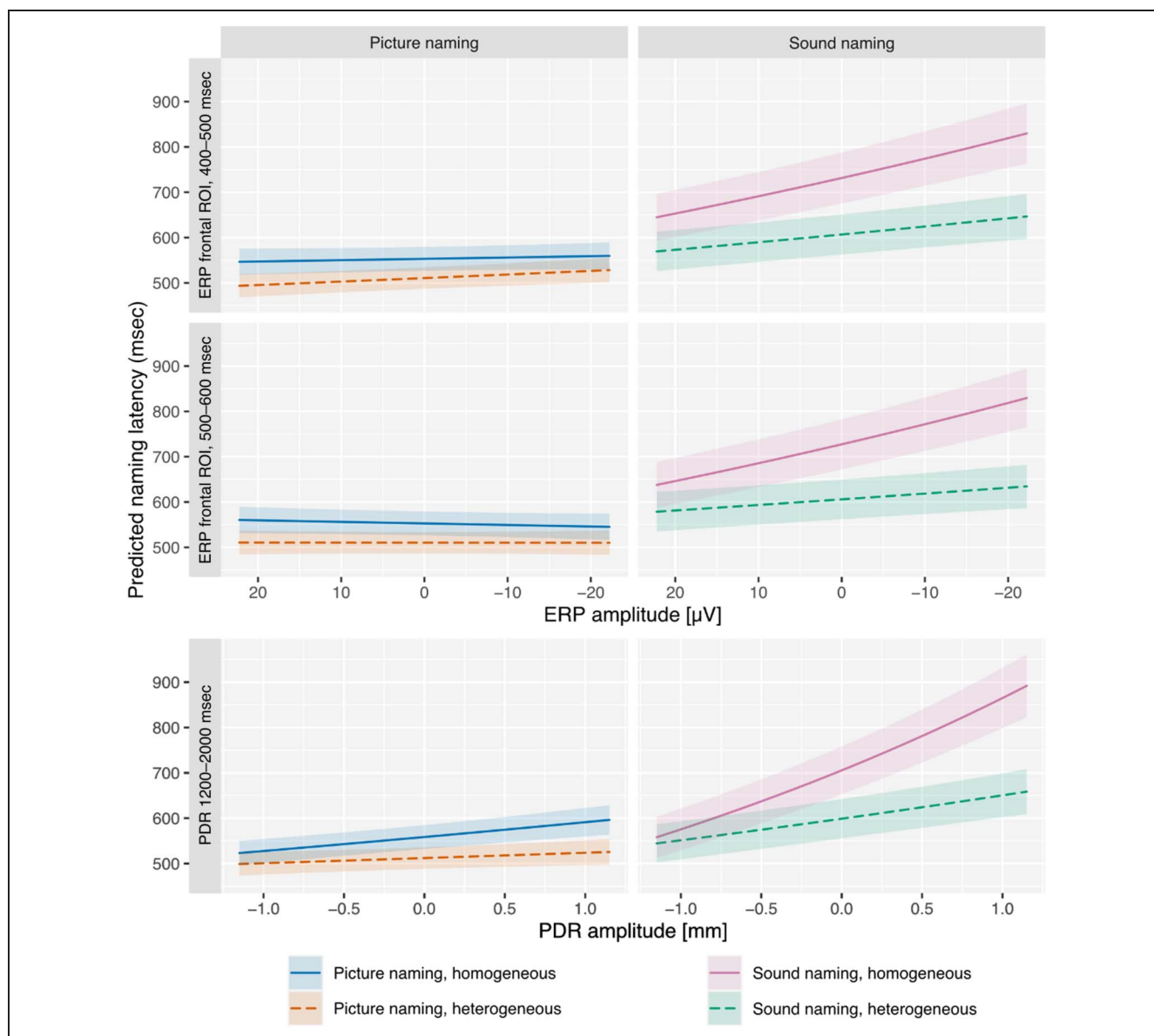
better than the model including only the main effects (model fit estimates are reported in Table 2). The naming latencies were slower in the homogeneous context than in the heterogeneous context for pictures (553 vs. 510 msec; Context effect for pictures 43 [29, 56] msec,  $t(32.8) = 6.27$ ;  $p < .001$ ) and also for sounds (735 vs. 606 msec; Context effect for sounds 129 [108, 150] msec,  $t(33.03) = 15$ ;  $p < .001$ ). The effect of Context was significantly larger for sounds than for pictures (Stimulus Type  $\times$  Context interaction effect  $-86$  [ $-99, -73$ ] msec,  $t(13120) = 15.41$ ,  $p < .001$ ; Stimulus Type effect in homogeneous context  $-182$  [ $-219, -144$ ] msec,  $t(31.36) = 12.2$ ;  $p < .001$ ;

Stimulus Type effect in heterogeneous context  $-96$  [ $-125, -66$ ] msec,  $t(31.28) = 7.38$ ;  $p < .001$ ).

Next, we report the LMM analyses including ERP or PDR amplitude as an additional predictor. With respect to the ERP amplitude, we focus on the two later time windows (400–500 msec and 500–600 msec; frontal ROI) because, in the two early time windows, there was no increase of the naming latency effects with an increase of the ERPs effects. The linear trends predicting the change in naming latency by ERP or PDR amplitude at mean ERP/PDR amplitude per context and stimulus type are illustrated in Figure 4.



**Figure 3.** (A) Mean naming latencies in msec for pictures and sounds by context and cycle (based on participant means). (B) Mean differences in naming latencies between homogeneous and heterogeneous contexts for pictures and sounds by cycle and difference of differences between picture and sound naming (reflecting the Stimulus Type  $\times$  Context interaction effect). Shaded areas reflect 95% CIs.



**Figure 4.** Relation between predicted naming latencies in msec and mean ERP amplitudes in  $\mu\text{V}$  and PDR amplitudes in mm, respectively, by stimulus type and context. Shaded areas reflect 95% CIs.

#### ERP Time Window 400–500 msec

Including the ERP mean amplitudes of the time window from 400–500 msec in the frontal ROI in the model of the naming latencies significantly improved the model fit. Naming latencies increased significantly with more negative ERP amplitudes for pictures in the heterogeneous context, but not in the homogeneous context ( $-0.78$  [ $-1.3, -0.26$ ] vs.  $-0.29$  [ $-0.88, 0.29$ ] msec/ $\mu\text{V}$ ; Context  $\times$  ERP<sub>400–500 msec</sub> interaction effect for pictures  $0.49$  [ $-0.3, 1.27$ ] msec/ $\mu\text{V}$ ,  $t(13161) = 1.33$ ,  $p = .183$ ) and for sounds in both the heterogeneous context and the homogeneous context, with a significantly stronger linear trend in the latter ( $-1.73$  [ $-2.38, -1.08$ ] vs.  $-4.15$  [ $-4.97, -3.33$ ] msec/ $\mu\text{V}$ ; Context  $\times$  ERP<sub>400–500 msec</sub> interaction effect for sounds  $-2.41$  [ $-3.41, -1.41$ ] msec/ $\mu\text{V}$ ,  $t(13090) = 3.78$ ,  $p < .001$ ). The difference in

the linear trend between heterogeneous and homogeneous contexts was significantly stronger for sounds than for pictures (Stimulus Type  $\times$  Context  $\times$  ERP<sub>400–500 msec</sub> interaction effect  $2.9$  [ $1.64, 4.16$ ] msec/ $\mu\text{V}$ ,  $t(13170) = 3.63$ ,  $p < .001$ ).

#### ERP Time Window 500–600 msec

Including the ERP mean amplitudes of the time window from 500–600 msec in the frontal ROI in the model of the naming latencies significantly improved the model fit. Naming latencies were not significantly related to ERP amplitudes for pictures in either the heterogeneous context or the homogeneous context ( $0.01$  [ $-0.5, 0.52$ ] vs.  $0.34$  [ $-0.21, 0.89$ ] msec/ $\mu\text{V}$ ; Context  $\times$  ERP<sub>500–600 msec</sub> interaction effect for pictures  $0.33$  [ $-0.41, 1.08$ ] msec/ $\mu\text{V}$ ,



$t(13150) = 0.84, p = .401$ ), but increased significantly with more negative ERP amplitudes for sounds in both the heterogeneous context and the homogeneous context, with a significantly stronger linear trend in the latter ( $-1.26 [-1.85, -0.66]$  vs.  $-4.3 [-5.08, -3.52]$  msec/ $\mu$ V; Context  $\times$  ERP<sub>500–600 msec</sub> interaction effect for sounds  $-3.04 [-3.99, -2.1]$  msec/ $\mu$ V,  $t(13110) = 5.49, p < .001$ ). The difference in the linear trend between heterogeneous and homogeneous contexts was significantly stronger for sounds than for pictures (Stimulus Type  $\times$  Context  $\times$  ERP<sub>500–600 msec</sub> interaction effect  $3.37 [2.17, 4.57]$  msec/ $\mu$ V,  $t(13170) = 4.46, p < .001$ ).

### PDR

Including the PDR mean amplitudes of the time window from 1200–2000 msec in the model of the naming latencies significantly improved the model fit. Naming latencies increased significantly with increasing PDR amplitudes for pictures in both the heterogeneous context and the homogeneous context, with a significantly stronger linear trend in the latter ( $11.5 [0.9, 22.2]$  vs.  $31.7 [19.9, 43.6]$  msec/mm; Context  $\times$  PDR interaction effect for pictures  $20.2 [4.6, 35.8]$  msec/mm,  $t(12394) = 2.34, p = .02$ ). Naming latencies increased significantly with increasing PDR amplitudes also for sounds in both the heterogeneous context and the homogeneous context, again with a significantly stronger linear trend in the latter ( $49.5 [36.2, 62.8]$  vs.  $143.9 [124.6, 163.1]$  msec/mm; Context  $\times$  PDR interaction effect for sounds  $94.3 [73, 115.7]$  msec/mm,  $t(12070) = 7.604, p < .001$ ). The difference in the linear trend between heterogeneous and homogeneous contexts was significantly stronger for sounds than for pictures (Stimulus Type  $\times$  Context  $\times$  PDR interaction effect  $-74.1 [-100, -48.2]$  msec/mm,  $t(12643) = -4.04, p < .001$ ).

## DISCUSSION

BCN is a prominent task in speech production research. The core finding of interference in a semantically homogeneous context compared with a semantically heterogeneous context is thought to be informative with respect to the mechanisms underlying lexical selection in word production. Nearly all BCN studies have used pictures to elicit naming responses, but a recent study by Wöhner and colleagues (2021) demonstrated that such interference is also obtained with sounds. Importantly, Wöhner and colleagues found that the magnitude of semantic interference was much larger for sounds than for pictures. The aim of the present study was to identify possible reasons of this differential pattern. To this end, we replicated Experiment 2 from Wöhner and colleagues while recording EEG and pupil diameter data in addition to naming latencies. Our key findings can be summarized as follows. The behavioral data replicated Wöhner and colleagues. Naming latencies were slower in the homogeneous context than in the heterogeneous context, and this slowing

was much larger for sounds than for pictures. The ERP data showed more positive amplitudes in the homogeneous compared with the heterogeneous context over central electrode locations between 140 and 180 msec for pictures and between 250 and 350 msec for pictures and sounds; in the latter time window, the effect was of the same magnitude for the two types of stimuli. The ERP data also showed a frontal negativity in the homogeneous compared with the heterogeneous context, which was larger for sounds than for pictures between 400 and 500 msec and only present for sounds between 500 and 600 msec. In the analysis of the pupillometric data, we observed a stronger pupil dilation in the homogeneous context compared with the heterogeneous context for sounds, but not for pictures. More negative ERP amplitudes in the 400- to 500-msec and 500- to 600-msec time windows (but not the 140- to 180-msec and 250- to 350-msec time windows) as well as higher PDR amplitudes predicted longer naming latencies for sound naming. The relation was moderated by context. That is, it was present in both contexts but much stronger in the homogeneous than in the heterogeneous context. These results are discussed in detail below.

## Behavioral Results

When the data from Cycle 1 were excluded from the analysis, semantic interference with pictures amounted to 43 msec ( $SE = 7$ ) in our study and to 38 msec ( $SE = 5$ ) in Wöhner and colleagues' study, respectively. Semantic interference with sounds was much larger, amounting to 129 msec ( $SE = 10$ ) in our study and to 129 msec ( $SE = 10$ ) in Wöhner and colleagues' study. Thus, the results from the analysis of naming latencies fully replicate the findings of Wöhner and colleagues (2021; Experiment 2).

## ERP Results

### Early Time Windows: 140–180 msec and 250–350 msec

At the central ROI, there was an enlarged positivity in the homogeneous context compared with the heterogeneous context for picture naming. At temporoparietal and mastoid ROIs, the polarity of the effect was reversed. This pattern replicates previous findings (Lin et al., 2022; Wang et al., 2018; Aristei et al., 2011) and is in line with the notion that the two effects reflect a dipolar component structure highlighted by average referencing. For pictures, the effect was present in the two early time windows (140–180 msec and 250–350 msec), which is in line with the notion that both the semantic level and the lexical level are involved in semantic interference in BCN (Belke, 2013). For sounds, the effect was only present in the time window from 250 to 350 msec. At first sight, one may take this as evidence that semantic and lexical processing is not (or to a much lesser extent) involved in sound naming. However, as we pointed out in the Introduction, ERP

effects for sounds could be delayed, as sound naming generally takes longer than picture naming (see Wöhner et al., 2021; Mädebach et al., 2017; see also the present behavioral data). One reason for this could be that, for sounds, the acoustic signal needs to unfold for some time before enough perceptual information is available for its identification, potentially leading to a delay in semantic and subsequent lexical processing. For pictures, on the other hand, all perceptual information is available simultaneously, which may allow for faster identification (for reviews, see Potter, 2014; De Lucia, Clarke, & Murray, 2010; Murray & Spierer, 2009). Regardless of such possible modality-dependent shifts in the timing of the ERP effects, our results unequivocally show that the ERP effects in the early time windows were not larger for sound naming than for picture naming. This finding is in clear contrast to what we have predicted. We have hypothesized that the larger context effect in the behavioral data for sound naming as compared with picture naming is because of differences in the perceptual and semantic processing of pictures and sounds (such as a more diffuse semantic activation induced by sounds as compared with pictures, because of coarse-to-fine semantic processing of sounds; see Murray et al., 2006, 2008). The absence of a larger context effect for sounds in the ERP amplitude in the early time windows calls this hypothesis into question. Rather, our ERP data suggest that the difference in semantic interference in picture and sound naming arises at a different, later processing stage.

#### *Later Time Windows: 400–500 msec and 500–600 msec*

In the time window from 400 to 500 msec, we observed a stronger negativity in the homogeneous context than in the heterogeneous context for both picture naming and sound naming, but the effect was much stronger for sound naming. In the time window from 500 to 600 msec, the effect was present for sound naming only. The results from LMM analyses that included ERP amplitude as an additional predictor suggest that this frontal negative component is a sensitive and specific marker of the enhanced context effect in sound naming as compared with picture naming: More negative ERP amplitudes predicted slower naming responses in the homogeneous context for sounds but not for pictures. The timing of this ERP effect suggests that late cognitive processes close to articulation are causing the differential behavioral pattern.

#### **Pupil Dilation Data**

We observed pupil constriction for pictures and pupil dilation for sounds. Moreover, for sounds, pupil dilation was larger in the homogeneous context than in the heterogeneous context, but no such effect was found for pictures. We briefly discuss these two findings.

As to the first finding, pupils respond to changes in luminance, contrast, and color (Carle, James, & Maddess, 2013;

Gamlin & McDougal, 2010). Specifically, the pupil constricts in response to changes of retinal illuminance (pupil light response; Mathôt, 2018; note that a transient constriction is also observed in response to the onset of stimuli darker than background; for a detailed discussion, see, Korn & Bach, 2016). Pupils not only respond to changes in illuminance but typically also dilate during the performance of cognitive tasks (such as naming) with more demanding or arousing tasks inducing larger pupil dilation. This task-evoked pupillary response (for a review, see, Mathôt, 2018; Beatty & Lucero-Wagoner, 2000) may already start at or even before stimulus onset as an anticipatory response. In our study, the task-evoked pupil dilation is expected to occur with both pictures and sounds. Presumably because of the temporal overlap of the illuminance-related pupil constriction and the task-related pupil dilation, we observed only a short pupil dilation at stimulus onset (and only a small pupil light response) with pictures, whereas we observed the expected task-related pupil dilation with sounds (without illuminance change).

Turning to the second finding, in their literature review, van der Wel and van Steenbergen (2018) proposed that pupil dilation is related to higher task demands and can be interpreted as an indirect indicator of effort in cognitive control tasks. In line with that, Mathôt (2018) suggested that the pupil dilates in response to stronger arousal and mental effort, as, for example, evidenced by early findings by Hess and Polt (1964) or Kahneman and Beatty (1966). In the homogeneous context, which potentially induces higher task demands because of semantic interference, a larger task-evoked pupil dilation could be expected compared with the heterogeneous context. Therefore, the question arises as to why such an increased pupil dilation in the homogeneous context was only observed for sounds but not for pictures, whereas semantic interference effects were observed at the behavioral level for both sounds and pictures. Provided that task-evoked pupillary response and pupil light response are noninteracting processes, it should have been possible to detect a greater pupil dilation in the homogeneous context during picture naming as well. As a parsimonious and plausible explanation for the observed pattern of results, we suggest that some cognitive process or competition for limited resources modulated by semantic interference might be involved only (or to a much larger extent) in sound but not in picture naming. The enhanced pupil dilation as well as the frontal negativity in the ERP in the late time window in the homogeneous compared with the heterogeneous context, both observed only during sound naming, might reflect sensitive and specific markers for such a cognitive process or resource competition, which is presumably responsible for the enhanced semantic interference effect in sound naming observed at the behavioral level.

Which cognitive processes or resources that are possibly involved in sound naming but not (or to a lesser extent)

in picture naming could be responsible for the context effects in the late ERP time windows and in the pupil data? Any such process should be sensitive to semantic information to account for the enhancement of the semantic context effect in sound naming. A possible candidate is self-monitoring. Self-monitoring acts as a cognitive control process in speech production and is used to inspect both prepared speech (internal self-monitoring) and produced speech (external self-monitoring) for (pragmatic, semantic, syntactic, or phonological) well-formedness (e.g., Lind & Hartsuiker, 2020; Levelt et al., 1999; Levelt, 1989; Baars, Motley, & Mackay, 1975). In their meta-analysis of behavioral and physiological word production studies, Indefrey and Levelt (2004) identified the bilateral superior temporal gyrus (STG) as brain regions that are involved in internal self-monitoring. Interestingly, two BCN studies with picture stimuli found effects of context in these areas. An intracerebral EEG study by Anders and colleagues (2019) found reduced activation in the homogeneous context in the right posterior STG. In addition, an fMRI-study by Hocking, McMahon, and de Zubicaray (2009) found greater activation in the left middle to posterior STG in the homogeneous context. These studies thus provide some evidence that self-monitoring processes might be modulated in BCN. Notably, the STG is relevant for auditory processing as well (the auditory cortex is located in the supratemporal plane and the posterior part of the STG; Rivier & Clarke, 1997; Celesia, 1976). Thus, parallel recruitment of auditory processing resources may occur during sound naming, when the processing of the auditory stimulus and verbal response preparation occur simultaneously, possibly interfering with self-monitoring. Given that pictures are processed in largely different brain areas, starting with the visual cortex, no such double recruitment of processing resources should occur during verbal response preparation with visual stimuli.

Indefrey and Levelt (2004) suggested that self-monitoring starts as soon as the word-initial phonological syllable has been composed and continues until after articulation. For a mean naming latency of 600 msec, Indefrey and Levelt estimated as earliest possible onset of self-monitoring at 355 msec after picture onset. The mean naming latency for sounds in the heterogeneous context was 606 msec in our study and closely matches the 600-msec reference from Levelt and Indefrey. Therefore, the long-lasting negativity starting at about 400 msec in the homogeneous context in sound naming substantially overlaps with the time window in which self-monitoring is assumed to take place.

The self-monitoring account finds further support by a study by Fargier and Laganaro (2019). In this study, participants named pictures while either paying attention to auditory syllables presented at different (positive) stimulus onset asynchronies or ignoring them or hearing no auditory syllables. Processing the (phonologically unrelated) auditory stimuli interfered with the naming process, as indexed by longer naming latencies. Related ERP

modulations were found in a time window about 200 msec before articulation. The authors interpreted the interference effect as resulting from attentional load and a competition for shared neural resources between the later stages in speech planning, in particular phonological encoding and self-monitoring, and processing of the auditory input. Our own data suggest that such interference of auditory processing on speech planning is not limited to auditory syllables but generalizes to nonverbal auditory stimuli such as environmental sounds. Concurrent sound processing and speech planning leads to competition for shared neural resources with higher effort for response preparation in sound naming. Because of the longer sound-naming latencies because of semantic interference in the homogeneous compared with the heterogeneous condition, the temporal overlap and resource competition between continued auditory stimulus processing and speech planning is extended, presumably amplifying the effects of semantic interference originating at earlier stages of processing. This higher effort is reflected in increased naming latencies, stronger pupil dilation, and larger ERP amplitudes from 400 msec in the homogeneous sound-naming context. Note, however, that this account should be considered speculative and tentative. Clearly, it needs independent validation. For example, the present data leave open whether temporal overlap in the processing of the continued input (processing of the sound stimuli) and the output (in particular phonological encoding and self-monitoring) as in the present experiment contributes to the observed enhanced semantic interference in sound naming as compared with picture naming or, alternatively, close temporal proximity could be sufficient. One way to disentangle these two possibilities would be to present the stimuli only briefly, so that they can be unambiguously identified, but the stimulation ends before participants engage in later production processes, such as phonological encoding and self-monitoring. In any case, further specification of the cognitive processes or shared neural resources responsible for the enhanced semantic interference in sound naming in future studies could significantly improve our understanding of speech production.

## Conclusion

We observed increased semantic interference in blocked-cyclic sound naming when compared with blocked-cyclic picture naming, replicating Wöhner and colleagues (2021). This difference was associated with a late frontal negativity in the ERP and stronger pupil dilation. Our data thus provide new insights into the neural and physiological correlates of semantic interference in BCN. Specifically, our findings indicate that increased semantic interference in sound naming does not arise during early perceptual or semantic processing, but rather during processes closer to articulation. We suggest that the processing of sounds competes for auditory resources with phonological

response preparation and self-monitoring, leading to enhanced semantic interference. From a more general perspective, the present study shows that expanding the toolbox for studying speech production beyond picture

naming (here, to include sound naming) and using a multi-method approach (here, to add ERPs and pupil dilation as additional dependent variables to naming latencies) brings new insights to the field.

## APPENDIX

**Table A1.** Target Names Used in the Experiment

<i>Set</i>	<i>Heterogeneous 1 (Mixed)</i>	<i>Heterogeneous 2 (Mixed)</i>	<i>Heterogeneous 3 (Mixed)</i>	<i>Heterogeneous 4 (Mixed)</i>
Homogeneous 1 (vehicles)	Auto [car]	Zug [train]	Schiff [ship]	Motorrad [motorcycle]
Homogeneous 2 (musical instruments)	Orgel [organ]	Dudelsack [bagpipe]	Gitarre [guitar]	Trommel [drum]
Homogeneous 3 (animals)	Schwein [pig]	Kuh [cow]	Pferd [horse]	Hund [dog]
Homogeneous 4 (tools)	Hammer [hammer]	Bohrer [drill]	Säge [saw]	Schere [scissors]

English translations appear in brackets.

**Table A2.** Model-implied ERP and PDR Amplitudes (per Time Window and ROI) and RTs per Context and/or Stimulus Type from the Preferred Model Including 95% CIs

<i>Dependent Variable and Context</i>	<i>Stimulus Type</i>		<i>Interaction (or Mean and Difference)</i>
	<i>Picture</i>	<i>Sound</i>	
ERP 140–180 msec, temporoparietal ROI [ $\mu$ V]			
Homogeneous	−0.55 [−1.49, 0.39]	−1.46 [−2.04, −0.89]	
Heterogeneous	−0.24 [−1.2, 0.73]	−1.41 [−1.96, −0.73]	Interaction
Difference	−0.31 [−0.55, −0.08]	−0.05 [−0.29, 0.19]	−0.26 [−0.48, −0.04]
ERP 140–180 msec, central ROI [ $\mu$ V]			
Homogeneous	0.74 [−0.14, 1.62]	3.06 [2.08, 4.05]	
Heterogeneous	0.18 [−0.73, 1.09]	2.93 [2, 3.86]	Interaction
Difference	0.56 [0.25, 0.88]	0.13 [−0.19, 0.45]	0.43 [0.07, 0.8]
ERP 250–350 msec, mastoidal ROI [ $\mu$ V]			
Homogeneous			−1.62 [−2.29, −0.96]
Heterogeneous			−1.03 [−1.63, −0.44]
Mean and Context effect	−0.7 [−1.47, 0.07]	−1.96 [−2.65, −1.27]	−0.59 [−0.8, −0.38]
ERP 250–350 msec, central ROI [ $\mu$ V]			
Homogeneous			0.7 [−0.04, 1.4]
Heterogeneous			0.09 [−0.57, 0.75]
Mean and Context effect	−1.08 [−2.04, −0.13]	1.87 [1.13, 2.61]	0.61 [0.35, 0.86]



**Table A2.** (continued)

Dependent Variable and Context	Stimulus Type		Interaction (or Mean and Difference)
	Picture	Sound	
ERP 400–500 msec, frontal ROI [ $\mu$ V]			
Homogeneous	–1.59 [–2.25, –0.93]	–2.31 [–3.25, –1.38]	
Heterogeneous	–1.23 [–1.79, –0.67]	–1.08 [–1.86, –0.31]	Interaction
Difference	–0.36 [–0.8, –0.08]	–1.23 [–1.67, –0.79]	0.87 [0.41, 1.33]
ERP 500–600 msec, frontal ROI [ $\mu$ V]			
Homogeneous	0.34 [–0.26, 0.94]	–2.16 [–3.11, –1.21]	
Heterogeneous	0.43 [–0.12, 0.98]	–0.53 [–1.32, 0.26]	Interaction
Difference	–0.09 [–0.58, 0.4]	–1.63 [–2.12, –1.14]	1.54 [1.05, 2.03]
PDR 1200–2000 msec [mm]			
Homogeneous	–0.1 [–0.16, –0.04]	0.28 [0.23, 0.33]	
Heterogeneous	–0.08 [–0.13, –0.02]	0.22 [0.17, 0.27]	Interaction
Difference	–0.02 [–0.04, 0.001]	0.06 [0.04, 0.08]	–0.07 [–0.1, –0.05]
Reaction times [msec]			
Homogeneous	553 [527, 580]	735 [677, 792]	
Heterogeneous	510 [487, 534]	606 [561, 651]	Interaction
Difference	43 [29, 56]	129 [108, 150]	–86 [–99, –73]

### Acknowledgments

We thank Rasha Abdel Rahman for the helpful discussion of the data. Our research was funded by Leipzig University.

Corresponding authors: Andreas Widmann, Wilhelm Wundt Institute for Psychology, Leipzig University, Neumarkt 9-19, Leipzig, Germany, 04109, e-mail: widmann@uni-leipzig.de or Jörg D. Jescheniak, Wilhelm Wundt Institute for Psychology, Leipzig University, Neumarkt 9-19, Leipzig, Germany, 04109, e-mail: jdj@uni-leipzig.de.

### Data Availability Statement

The stimuli, the presentation lists, the trial-level behavioral and ERP and PDR time window mean amplitude data, and the statistical analysis scripts are openly available via the Open Science Framework: <https://osf.io/9hjtq/>. The ERP and pupil raw data are available upon request (e-mail to widmann@uni-leipzig.de).

### Author Contributions

Magdalena Gruner: Conceptualization; Data curation; Formal analysis; Investigation; Writing—Original draft;

Writing—Review & editing. Andreas Widmann: Conceptualization; Data curation; Formal analysis; Methodology; Software; Supervision; Writing—Original draft; Writing—Review & editing. Stefan Wöhner: Conceptualization; Resources; Writing—Review & editing. Erich Schröger: Conceptualization; Writing—Review & editing. Jörg D. Jescheniak: Conceptualization; Funding acquisition; Supervision; Writing—Original draft; Writing—Review & editing.

### Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were  $M(\text{an})/M = .407$ ,  $W(\text{oman})/M = .32$ ,  $M/W = .115$ , and  $W/W = .159$ , the comparable proportions for the articles that these authorship teams cited were  $M/M = .549$ ,  $W/M = .257$ ,  $M/W = .109$ , and  $W/W = .085$  (Postle and

Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

## Note

1. Note that the overall pattern in the literature is not fully consistent because some studies reported a stronger negativity in the homogeneous context compared with the heterogeneous context at frontal or rather anterior electrodes (e.g., Python et al., 2018; Janssen et al., 2015).

## REFERENCES

- Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 604–614. <https://doi.org/10.1037/0278-7393.33.3.604>, PubMed: 17470008
- Abdel Rahman, R., & Melinger, A. (2011). The dynamic microstructure of speech production: Semantic interference built on the fly. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 149–161. <https://doi.org/10.1037/a0021208>, PubMed: 20854006
- Anders, R., Llorens, A., Dubarry, A.-S., Trébuchon, A., Liégeois-Chauvel, C., & Alario, F.-X. (2019). Cortical dynamics of semantic priming and interference during word production: An intracerebral study. *Journal of Cognitive Neuroscience*, 31, 978–1001. [https://doi.org/10.1162/jocn\\_a\\_01406](https://doi.org/10.1162/jocn_a_01406), PubMed: 30938588
- Aristei, S., Melinger, A., & Abdel Rahman, R. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, 23, 1567–1586. <https://doi.org/10.1162/jocn.2010.21474>, PubMed: 20515409
- Baars, B. J., Motley, M. T., & Mackay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14, 382–391. [https://doi.org/10.1016/S0022-5371\(75\)80017-X](https://doi.org/10.1016/S0022-5371(75)80017-X)
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3, 12–28. <https://doi.org/10.21500/20112084.807>
- Barthel, M., & Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science*, 43, e12768. <https://doi.org/10.1111/cogs.12768>, PubMed: 31310021
- Bartón, K. (2023). *MuMIn: Multi-Model Inference* (R-package, version 1.47.5) [Computer software]. <https://cran.r-project.org/package=MuMIn>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinari, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 142–162). Cambridge University Press.
- Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language*, 69, 228–256. <https://doi.org/10.1016/j.jml.2013.05.008>
- Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical-semantic competition hypothesis. *Cognition*, 96, B45–B54. <https://doi.org/10.1016/j.cognition.2004.11.006>, PubMed: 15925568
- Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *Quarterly Journal of Experimental Psychology*, 58, 667–692. <https://doi.org/10.1080/02724980443000142>
- Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *Quarterly Journal of Experimental Psychology*, 66, 2135–2160. <https://doi.org/10.1080/17470218.2013.775318>, PubMed: 23521507
- Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review*, 3, 395–421. <https://doi.org/10.3758/BF03214545>, PubMed: 24213975
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. <https://doi.org/10.1163/156856897X00357>, PubMed: 9176952
- Brownsett, S. L., Mascelloni, M., Gowlett, G., McMahon, K. L., & de Zubicaray, G. I. (2022). Neighing dogs: Semantic context effects of environmental sounds in spoken word production—A replication and extension. *Quarterly Journal of Experimental Psychology*, 76, 1990–2000. <https://doi.org/10.1177/17470218221137007>, PubMed: 36301012
- Carle, C. F., James, A. C., & Maddess, T. (2013). The pupillary response to color and luminance variant multifocal stimuli. *Investigative Ophthalmology & Visual Science*, 54, 467–475. <https://doi.org/10.1167/iovs.12-10829>, PubMed: 23188728
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind and Content*, 41, 63–65. <https://doi.org/10.1093/mind/os-XI.41.63>
- Celesia, G. G. (1976). Organization of auditory cortical areas in man. *Brain*, 99, 403–414. <https://doi.org/10.1093/brain/99.3.403>, PubMed: 1000279
- Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, 114, 389–404. <https://doi.org/10.1016/j.cognition.2009.10.012>, PubMed: 19909945
- Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1554–1568. <https://doi.org/10.1037/a0024329>, PubMed: 21688942
- Clarke, A. (2015). Dynamic information processing states revealed through neurocognitive models of object semantics. *Language, Cognition and Neuroscience*, 30, 409–419. <https://doi.org/10.1080/23273798.2014.970652>, PubMed: 25745632
- Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1372–1384. <https://doi.org/10.1037/0278-7393.31.6.1372>, PubMed: 16393052
- Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, 81, B77–B86. [https://doi.org/10.1016/S0010-0277\(01\)00135-4](https://doi.org/10.1016/S0010-0277(01)00135-4), PubMed: 11483172
- De Lucia, M., Clarke, S., & Murray, M. M. (2010). A temporal hierarchy for conspecific vocalization discrimination in humans. *Journal of Neuroscience*, 30, 11210–11221. <https://doi.org/10.1523/JNEUROSCI.2239-10.2010>, PubMed: 20720129
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321. <https://doi.org/10.1037/0033-295X.93.3.283>, PubMed: 3749399

- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>, PubMed: 15102499
- Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One*, *7*, e30135. <https://doi.org/10.1371/journal.pone.0030135>, PubMed: 22355308
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*, 243. <https://doi.org/10.3389/fpsyg.2011.00243>, PubMed: 22007180
- Fargier, R., & Laganaro, M. (2019). Interference in speaking while hearing and vice versa. *Scientific Reports*, *9*, 5375. <https://doi.org/10.1038/s41598-019-41752-7>, PubMed: 30926840
- Feng, C., Damian, M. F., & Qu, Q. (2021). Parallel processing of semantics and phonology in spoken production: Evidence from blocked cyclic picture naming and EEG. *Journal of Cognitive Neuroscience*, *33*, 725–738. [https://doi.org/10.1162/jocn\\_a\\_01675](https://doi.org/10.1162/jocn_a_01675), PubMed: 33475451
- Gamlin, P. D. R., & McDougal, D. H. (2010). Pupil. In D. A. Dartt (Eds.), *Encyclopedia of the eye* (pp. 549–555). Academic Press. <https://doi.org/10.1016/B978-0-12-374203-2.00231-1>
- Hendrickson, K., Walenski, M., Friend, M., & Love, T. (2015). The organization of words and environmental sounds in memory. *Neuropsychologia*, *69*, 67–76. <https://doi.org/10.1016/j.neuropsychologia.2015.01.035>, PubMed: 25624059
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*, 1190–1192. <https://doi.org/10.1126/science.143.3611.1190>, PubMed: 17833905
- Hocking, J., McMahan, K. L., & de Zubicaray, G. I. (2009). Semantic context and visual feature effects in object naming: An fMRI study using arterial spin labeling. *Journal of Cognitive Neuroscience*, *21*, 1571–1583. <https://doi.org/10.1162/jocn.2009.21114>, PubMed: 18823254
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*, 464–482. <https://doi.org/10.1016/j.cognition.2005.02.006>, PubMed: 16413014
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 255. <https://doi.org/10.3389/fpsyg.2011.00255>, PubMed: 22016740
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>, PubMed: 15037128
- Janssen, N., Carreiras, M., & Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *Neuroimage*, *57*, 1243–1250. <https://doi.org/10.1016/j.neuroimage.2011.05.015>, PubMed: 21600993
- Janssen, N., Hernández-Cabrera, J. A., van der Meij, M., & Barber, H. A. (2015). Tracking the time course of competition during word production: Evidence for a post-retrieval mechanism of conflict resolution. *Cerebral Cortex*, *25*, 2960–2969. <https://doi.org/10.1093/cercor/bhu092>, PubMed: 24825785
- Jeon, H. A., & Lee, K. M. (2009). Diagnostic utility of sound naming in early Alzheimer's disease. *Journal of the International Neuropsychological Society*, *15*, 231–238. <https://doi.org/10.1017/S135561770909033X>, PubMed: 19203436
- Kadem, M., Herrmann, B., Rodd, J. M., & Johnsrude, I. S. (2020). Pupil dilation is sensitive to semantic ambiguity and acoustic degradation. *Trends in Hearing*, *24*, 1–16. <https://doi.org/10.1177/2331216520964068>, PubMed: 33124518
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585. <https://doi.org/10.1126/science.154.3756.1583>
- Kitazawa, Y., Sonoda, M., Sakakura, K., Mitsunashi, T., Firestone, E., Ueda, R., et al. (2023). Intra- and inter-hemispheric network dynamics supporting object recognition and speech production. *Neuroimage*, *15*, 119954. <https://doi.org/10.1016/j.neuroimage.2023.119954>, PubMed: 36828156
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1–16.
- Korn, C. W., & Bach, D. R. (2016). A solid frame for the window on cognition: Modeling event-related pupil responses. *Journal of Vision*, *16*, 28. <https://doi.org/10.1167/16.3.28>, PubMed: 26894512
- Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior Research Methods*, *51*, 1336–1342. <https://doi.org/10.3758/s13428-018-1075-y>, PubMed: 29992408
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., et al. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*, 23–34. <https://doi.org/10.1111/j.1469-8986.2012.01477.x>, PubMed: 23157603
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2020). *lmerTest: Tests in linear mixed effects models* (R package, version 3.1–3) [Computer software]. <https://cran.r-project.org/package=lmerTest>
- Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*, 1–14. <https://doi.org/10.1177/2515245920951503>
- Lakens, D., & Caldwell, A. (2022). *Superpower: Simulation-based power analysis for factorial designs* (R-package, version 0.2.0) [Computer software]. <https://CRAN.R-project.org/package=Superpower>
- Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means* (R-package, version 1.8.9) [Computer software]. <https://cran.r-project.org/package=emmeans>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/6393.001.0001>
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75. <https://doi.org/10.1017/S0140525X99001776>, PubMed: 11301520
- Lin, H. P., Kuhlen, A. K., Melinger, A., Aristei, S., & Abdel Rahman, R. (2022). Concurrent semantic priming and lexical interference for close semantic relations in blocked-cyclic picture naming: Electrophysiological signatures. *Psychophysiology*, *59*, e13990. <https://doi.org/10.1111/psyp.13990>, PubMed: 34931331
- Lind, A., & Hartsuiker, R. J. (2020). Self-monitoring in speech production: Comprehending the conflict between conflict- and comprehension-based accounts. *Journal of Cognition*, *3*, 16. <https://doi.org/10.5334/joc.118>, PubMed: 32944679
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. <https://doi.org/10.3389/fpsyg.2015.01171>, PubMed: 26300841
- Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*, 551–569. <https://doi.org/10.1037/a0039184>, PubMed: 25938178
- Mädebach, A., Kiesel, M. L., & Jescheniak, J. D. (2018). Localizing semantic interference from distractor sounds in picture naming: A dual-task study. *Psychonomic Bulletin & Review*, *25*, 1909–1916. <https://doi.org/10.3758/s13423-017-1386-5>, PubMed: 29030759



- Mädebach, A., Widmann, A., Posch, M., Schröger, E., & Jescheniak, J. D. (2022). Hearing “birch” hampers saying “duck”—An ERP study on phonological interference in immediate and delayed word production. *Journal of Cognitive Neuroscience*, *34*, 1397–1415. [https://doi.org/10.1162/jocn\\_a\\_01859](https://doi.org/10.1162/jocn_a_01859), PubMed: 35551402
- Mädebach, A., Wöhner, S., Kieseler, M.-L., & Jescheniak, J. D. (2017). Neighing, barking, and drumming horses—Object related sounds help and hinder picture naming. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1629–1646. <https://doi.org/10.1037/xhp0000415>, PubMed: 28414498
- Maess, B., Friederici, A. F., Damian, M. F., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, *14*, 455–462. <https://doi.org/10.1162/089892902317361967>, PubMed: 11970804
- Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*, 16. <https://doi.org/10.5334/joc.18>, PubMed: 31517190
- Merritt, S. L., Keegan, A. P., & Mercer, P. W. (1994). Artifact management in pupillometry. *Nursing Research*, *43*, 56–59. <https://doi.org/10.1097/00006199-199401000-00012>, PubMed: 8295842
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, *14*, 452–465. <https://doi.org/10.1093/cercor/bhh007>, PubMed: 15028649
- Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *Journal of Neuroscience*, *26*, 1293–1302. <https://doi.org/10.1523/JNEUROSCI.4511-05.2006>, PubMed: 16436617
- Murray, M. M., Camen, C., Spierer, L., & Clarke, S. (2008). Plasticity in representations of environmental sounds revealed by electrical neuroimaging. *Neuroimage*, *39*, 847–856. <https://doi.org/10.1016/j.neuroimage.2007.09.002>, PubMed: 17950001
- Murray, M. M., & Spierer, L. (2009). Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. *Hearing Research*, *258*, 121–133. <https://doi.org/10.1016/j.heares.2009.04.022>, PubMed: 19454307
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227–252. <https://doi.org/10.1016/j.cognition.2009.09.007>, PubMed: 19854436
- Ouyang, G., Sommer, W., & Zhou, C. (2015). A toolbox for residue iteration decomposition (RIDE)—A method for the decomposition, reconstruction, and single trial analysis of event related potentials. *Journal of Neuroscience Methods*, *250*, 7–21. <https://doi.org/10.1016/j.jneumeth.2014.10.009>, PubMed: 25455337
- Papesh, M. H., & Goldinger, S. D. (2012). Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics*, *74*, 754–765. <https://doi.org/10.3758/s13414-011-0263-y>, PubMed: 22231605
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. <https://doi.org/10.1163/156856897X00366>, PubMed: 9176953
- Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181–197. <https://doi.org/10.1016/j.neuroimage.2019.05.026>, PubMed: 31103785
- Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology*, *60*, 211–229. <https://doi.org/10.1080/17470210600673818>, PubMed: 17455055
- Potter, M. C. (2014). Detecting and remembering briefly presented pictures. In K. Kveraga & M. Bar (Eds.), *Scene vision: Making sense of what we see* (pp. 177–197). Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/9780262027854.003.0009>
- Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859–862. <https://doi.org/10.3758/BF03192979>, PubMed: 18183901
- Python, G., Fargier, R., & Laganaro, M. (2018). ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex*, *99*, 1–12. <https://doi.org/10.1016/j.cortex.2017.09.008>, PubMed: 29121484
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org>
- Riley, E., McMahon, K. L., & de Zubicaray, G. (2015). Long-lasting semantic interference effects in object naming are not necessarily conceptually mediated. *Frontiers in Psychology*, *6*, 578. <https://doi.org/10.3389/fpsyg.2015.00578>, PubMed: 25999892
- Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *Neuroimage*, *6*, 288–304. <https://doi.org/10.1006/nimg.1997.0304>, PubMed: 9417972
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142. [https://doi.org/10.1016/0010-0277\(92\)90041-F](https://doi.org/10.1016/0010-0277(92)90041-F), PubMed: 1582154
- Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, *172*, 59–72. <https://doi.org/10.1016/j.cognition.2017.12.007>, PubMed: 29232595
- Schmidtke, J. (2018). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, *40*, 529–549. <https://doi.org/10.1017/S0272263117000195>
- Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology*, *55*, 121–132. <https://doi.org/10.1027/1618-3169.55.2.121>, PubMed: 18444522
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture–word interference studies. *Journal of Memory and Language*, *29*, 86–102. [https://doi.org/10.1016/0749-596X\(90\)90011-N](https://doi.org/10.1016/0749-596X(90)90011-N)
- Steinhauer, S. R., Bradley, M. M., Siegle, G. J., Roecklein, K. A., & Dix, A. (2022). Publication guidelines and recommendations for pupillary measurement in psychophysiological studies. *Psychophysiology*, *59*, e14035. <https://doi.org/10.1111/psyp.14035>, PubMed: 35318693
- Strauch, C., Wang, C.-A., Einhäuser, W., Van der Stigchel, S., & Naber, M. (2022). Pupillometry as an integrated readout of distinct attentional networks. *Trends in Neurosciences*, *45*, 635–647. <https://doi.org/10.1016/j.tins.2022.05.003>, PubMed: 35662511
- van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*, 584–589. <https://doi.org/10.3758/BF03193889>, PubMed: 17393828
- van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review.



- Psychonomic Bulletin & Review*, 25, 2005–2015. <https://doi.org/10.3758/s13423-018-1432-y>, PubMed: 29435963
- Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, 114, 830–841. <https://doi.org/10.1037/0033-295X.114.3.830>, PubMed: 17638508
- Wang, M., Shao, Z., Chen, Y., & Schiller, N. O. (2018). Neural correlates of spoken word production in semantic and phonological blocked cyclic naming. *Language, Cognition, and Neuroscience*, 33, 575–586. <https://doi.org/10.1080/23273798.2017.1395467>
- Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data—A practical approach. *Journal of Neuroscience Methods*, 30, 34–46. <https://doi.org/10.1016/j.jneumeth.2014.08.002>, PubMed: 25128257
- Wöhner, S., Jescheniak, J. D., & Mädebach, A. (2020). Semantic interference is not modality specific: Evidence from sound naming with distractor pictures. *Quarterly Journal of Experimental Psychology*, 73, 2290–2308. <https://doi.org/10.1177/1747021820943130>, PubMed: 32640868
- Wöhner, S., Luckow, J., Brandt, M., Stahlmann, J., Werwach, A., & Jescheniak, J. D. (2024). Semantic facilitation in blocked picture categorization: Some data and considerations regarding task selection. *Journal of Experimental Psychology: Human Perception and Performance*, 50, 515–530. <https://doi.org/10.1037/xhp0001203>, PubMed: 38546627
- Wöhner, S., Mädebach, A., & Jescheniak, J. D. (2021). Naming pictures and sounds: Stimulus type affects semantic context effects. *Journal of Experimental Psychology: Human Perception and Performance*, 47, 716–730. <https://doi.org/10.1037/xhp0000908>, PubMed: 34264729
- Wöhner, S., Mädebach, A., Schriefers, H., & Jescheniak, J. D. (2024). Adaptive lexical processing of semantic competitors extends to alternative names: Evidence from blocked-cyclic picture naming. *Quarterly Journal of Experimental Psychology*, 17470218241245107. <https://doi.org/10.1177/17470218241245107>, PubMed: 38514252
- Yao, D. (2001). A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiological Measurement*, 22, 693–711. <https://doi.org/10.1088/0967-3334/22/4/305>, PubMed: 38546627
- Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog's meow: Asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research*, 193, 603–614. <https://doi.org/10.1007/s00221-008-1664-6>, PubMed: 19066869