Different electrophysiological signatures of similarity-induced and Stroop-like interference in language production

Svetlana Pinet [a,b] and Nazbanou Nozari [c, d]

[a] Basque Center on Cognition, Brain, and Language (BCBL)

[b] Ikerbasque, Basque Foundation for Science

[c] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

[d] Center for the Neural Basis of Cognition (CNBC), Pittsburgh, PA, USA

Corresponding author: Nazbanou Nozari

Department of Psychology, Baker Hall 335H, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

Email address: bnozari@andrew.cmu.edu

Abstract

Contextual similarity between targets and competitors, whether semantic or phonological, often leads to behavioral interference in language production. It has been assumed that resolving such interference relies on control processes similar to those involved in tasks such as Stroop. This paper tests this assumption by comparing the electrophysiological signatures of interference resulting from a contextual similarity vs. a Stroop-like manipulation. In blocks containing two items, participants repeatedly named pictures that were semantically related, phonologically related or unrelated (*contextual similarity manipulation*). In straight blocks, the pictures were named by their canonical names. In reverse blocks, participants had to reverse the names (*Stroop-like manipulation*). Both manipulations led to behavioral interference, but with different electrophysiological profiles. Whole-scalp stimulus-locked and response-locked analyses of semantic and phonological similarity pointed to a system with global modularity with some degree of cascading and interactivity, whereas the effect of phase reversal was sustained and of the opposite polarity. More strikingly, a representational similarity analysis showed a biphasic pattern for Stroop-like reversal, with earlier higher similarity scores for the reverse phase flipping into lower scores ~500 ms post-stimulus onset. In contrast, contextual similarity induced higher similarity scores up to articulation. Finally, response-locked mediofrontal components indexing performance monitoring differed between manipulations. Correct response negativity's (CRN) amplitude was lower in the phonological blocks, whereas a pre-CRN component had higher amplitude in reverse vs. straight blocks. These results argue against the involvement of Stroop-like control mechanisms in resolving interference from contextual similarity in language production.

Keywords: language production; semantic interference; phonological interference; Stroop; inhibitory control; EEG; conflict detection; ERN/CRN

Different electrophysiological signatures of similarity-induced and Stroop-like interference in language production

One of the most robust effects in language production is interference from similar items; naming a picture of "cat" is more difficult in the context of a taxonomically related word like "dog" or phonologically related items like "mat" (e.g., Belke et al., 2005; Nozari et al., 2016). This effect has often been attributed to the competition between the target and its contextual competitor. Since competition is a hallmark of Stroop-like tasks, in which a distractor word competes with the name of a target, it is conceivable that similar mechanisms may underlie the interference induced by similarity and Stroop-like manipulations. Such an assumption, if true, has both theoretical and clinical implications. On the theoretical side, it would call for a unified model of inhibitory control in language production that explains similarity-induced and Stroop-like effects under the same mechanism. On the clinical side, it would predict inhibitory control abilities measured through Stroop-like tasks to be predictive of similarity-induced interference. This paper tests this basic assumption by comparing the electrophysiological signatures of two types of similarity (semantic and phonological) and a Stroop-like manipulation in language production within the same paradigm.

Signatures of contextual similarity in behavioral and EEG data

With the exception of the facilitatory effect of contexts that provide thematic relations between items (e.g., Alario et al., 2000; Mahon et al., 2007; McDonagh et al., 2020; Oppenheim & Nozari, 2021; cf., Abdel Rahman & Melinger, 2007), the bulk of studies investigating the effect of semantically related competitors on word production have reported interference, in the form of longer response times (RTs) and/or higher error rates (e.g., Belke et al., 2005; Costa et al., 2009; Schnur et al., 2006, 2009). Contextual interference is not limited to semantically related competitors. Phonologically related competitors also interfere with production. Participants take longer to name a picture like "cat" if it appears in the context of phonological related items like "mat" compared to unrelated items like "shoe" (Feng et al., 2021; Nozari et al., 2016, 2016; Qu et al., 2021). An exception is observed when the majority of items overlap in the onset (e.g., "cap"/"cat") which leads to facilitation (e.g., Nozari et al., 2016; Roelofs, 1999; Wang et al., 2018). But this facilitation effect has been shown to be strategic (Nozari et al.,

2016; O'Séaghdha & Frazer, 2014). In short, both semantic and phonological competitors generally induce a robust interference in the language production system.

Electrophysiologically, semantic interference has been associated with lower amplitude of ERP components in the 200-500ms post-stimulus time-window (Costa et al., 2009; Python et al., 2018b; Wong et al., 2017). For example, in the blocked cyclic naming paradigm, where participants repeatedly name a small set of pictures, a lower negative amplitude was seen across cycles for related vs. unrelated items up to 450 ms after the stimulus presentation (Janssen et al., 2011, 2015). Additionally, a lower P2 amplitude was observed in the related condition (Python et al., 2018b). In the continuous sequential naming paradigm, semantic similarity manifested as a graded decrease in N2 and a graded increase in P3 with ordinal position, which was positively correlated with naming latencies (Costa et al., 2009; but see Llorens et al., 2014, for a failure to replicate). Interestingly, semantic facilitation in a Picture-Word Interference (PWI) paradigm also led to a less negative amplitude in similar time-windows (Python et al., 2018a). Although the similarity between the electrophysiological signatures of semantic interference and semantic facilitation reported in the studies listed above is not always replicated (e.g., Lin et al., 2022), these reports have raised the question of whether such ERP differences truly reflect interference, or rather simply the degree of featural overlap in related conditions regardless of the behavioral facilitation or interference (Nozari & Pinet, 2020).

Electrophysiological evidence for phonological overlap closely mirrors that of semantic overlap: regardless of the behavioral outcome, phonological overlap leads to lower amplitudes in EEG data. This holds true even when semantic and phonological overlap have been studied together using the block-cycling naming paradigm. Wang et al. (2018) reported lower amplitudes in the 200-550 ms time window for semantic and 350 to 550ms time window for phonological effects, compared to unrelated contexts, associated with semantic interference and phonological (onset) facilitation, respectively. More recently, Feng et al. (2021) found a similar ERP pattern within the same time window (180-380ms range), but associated with semantic and phonological interference, behaviorally. To summarize, ERP data do not appear to clearly distinguish between facilitation and interference induced by semantic or phonological similarity; rather, any kind of representational overlap between the items, regardless of the behavioral outcome, leads to lower amplitudes in ERPs.

Mechanisms of contextual interference and link to inhibitory control

Generally speaking, two types of mechanisms have been proposed to explain contextual interference in language production. The first attributes contextual interference to a competitive selection mechanism. For example, in WEAVER (Roelofs, 1992 et passim), a word can only be selected if its activation exceeds that of its competitors by a certain amount (i.e., a relative threshold). If competition is too strong, the threshold cannot be reached quickly, and production is hampered. To help reach the relative threshold necessary for selection, the model implements an active suppression mechanism for all the non-target lexical items. This suppression process is akin to the deployment and application of inhibitory control, thus linking contextual interference to mechanisms of inhibitory control (see also Roelofs, 2018).

The second type of mechanism, often referred to as a non-competitive selection mechanism (e.g., Oppenheim et al., 2010) proposes that contextual interference is not a product of a relative threshold during selection, but rather of incremental learning across trials, according to the delta rule. For example, the target "cat" activates "dog" through shared features, such as "four legs". The selection of "cat" strengthens the connections between the lexical item "cat" and the shared feature, while simultaneously weakening the connection between "dog" and the same feature. In other words, the system tries to learn to be more efficient the next time it encounters "cat". The dark side of this learning is that the system will be less efficient if it instead sees "dog". Because of the weakened connections between "dog" and some of its semantic features, it will be less activated and may not quickly pass an absolute threshold necessary for selection. Recently, the same incremental learning mechanism has been proposed for contextual interference induced by phonological similarity (Breining et al., 2019; Qu et al., 2021). The role of inhibitory control is less clear in incremental learning vs. competitive selection models. In a simple simulation of RTs, Oppenheim et al. (2010) showed that a linear booster (which does not selectively inhibit competitors) was sufficient to simulate the empirical pattern. While this demonstration does not rule out the involvement of inhibitory control in resolving similarity-induced interference, it questions its importance and relevance.

Empirical evidence for a link between inhibitory control and contextual interference is sparse and mixed (and likely also subject to publication bias, as null effects are less likely to be published). Crowther & Martin (2014) reported that the increase in naming latencies across cycles in semantically related blocks of a blocked cyclic naming task

was correlated with Stroop interference, but it is unclear from the report whether that correlation survives correction for many comparisons presented in that study. Moreover, they did not report a test of the interference size, namely the difference between related and unrelated conditions. Using a PWI task, Korko et al. (2021) found a correlation between the magnitude of semantic interference in PWI and conflict in the Flanker task (in which participants must suppress the flanking visual stimuli in order to respond to a central stimulus), but not the Simon task (in which participants must resolve the conflict resulting from a mismatch in the position of the visual stimuli on the screen and the hand they respond with), or the anti-saccade task (in which participants must suppress the urge to make an automatic saccade towards an external stimulus appearing on the left or right side of the screen, and instead move their eyes in the opposite direction). In another series of studies, while general naming RTs were shown to correlate with stop-signal RT (Shao et al., 2012; Xue et al., 2008; but see Higby et al., 2019), the magnitude of semantic interference was not (Shao et al., 2013). Finally, Shao et al. (2015) showed that the slopes of interference for the longest naming RTs correlated with the magnitude of the mean semantic interference effect, but no such correlation was found between the slopes and the mean interference effect in the Stroop task. The authors concluded, counterintuitively, that the role of inhibitory control was greater in resolving semantic interference than in resolving Stroop interference.

To summarize, different accounts of contextual interference vary in how much they rely on inhibitory control. Empirical evidence for the link between contextual interference and inhibitory control is also mixed. Moreover, these data are often of a correlational nature, which makes them vulnerable to issues of test reliability (Strauss et al., 2005), task impurity (Miyake et al., 2000), and statistical power. This study proposes a novel way to examine this link.

The current study

Whether contextual interference relies on similar inhibitory control mechanisms as those observed in inhibitory control tasks remains an open question. Nozari et al. (2016) looked at this question in a simplified version of the blocked cycling naming paradigm with only two items per block. The task had two manipulations: a *contextual similarity manipulation*, and a *Stroop-like phase manipulation*. To examine the effect of contextual similarity, the relationship between the two items within a block was manipulated to create semantically related, phonologically related, and unrelated blocks. To create a Stroop-like effect, each block containing a pair of items was named

in two phases, a *straight phase* and a *reverse phase*. In the straight phase, participants named each picture by its canonical name. In the reverse phase, they had to name the other item in the block upon viewing each picture. For example, if a block contained pictures of cat and dog, in the reversed phase, upon being presented with the picture of cat, participants would say "dog". This name reversal created a Stroop-like effect without involving cognitive processes such as reading that are not routinely engaged during conversational speech. People rarely have to suppress the written form of a word that appears in front of them as an alternative to the word they had planned in the middle of a conversation. On the other hand, we do encounter situations where we mistakenly recall one word when we need another, especially when visual cues for the word we need are not present. Imagine trying to round up your kids to drive them to school. One child is standing in front of you. Their name is prepotent because of the present visual signal, but calling their name will not help you. You need the name of the absent child, for whom you do not have immediate visual cues. In fact, the real-life equivalent of Stroop-like tasks in everyday language production, which by definition requires word retrieval from memory, frequently involve differences in prepotency as a function of less present cues for the to-be-produced word, which in turn leads to its disadvantage in a high-conflict situation. Our Stroop-like phase manipulation captures this effect.

The Stroop-like manipulation elicits indirect competitive inhibition (e.g., Munakata et al., 2011), a mechanism by which conflict is resolved between two competing alternatives. Theoretically speaking, the same mechanism could be at work for resolving conflict between the target and related competitors in the manipulations of contextual similarity. Nozari et al. (2016) observed both semantic and phonological behavioral interference, as well as a Stroop-like cost for reversal. Importantly, the effects of contextual similarity and phase did not interact, hinting at potentially different underlying processes for resolving contextual interference and Stroop-like interference. It is, however, possible that behavioral measures were not sensitive enough to detect a common mechanism underlying both effects, namely a similar inhibitory control process. This study aims to remedy that by investigating the electrophysiological correlates of these two manipulations.

The current study is an upgrade of the design of Nozari et al. (2016) in three ways: (1) we used the same target item (e.g., "cake") paired with a semantically similar item (e.g., "pie"), a phonologically similar item (e.g., "rake"), and an unrelated item, so that

each target is its own control. (2) We collected both response times (RTs) and response durations, as interference may manifest in either or there may be a speed-accuracy tradeoff (Damian, 2003; Kello et al., 2000). (3) Finally, we collected EEG data while participants completed the task. On the behavioral side, we expected to replicate the contextual interference for both semantically and phonologically related items, as well as a robust interference induced by reversing the names. As before, we expected no interactions between the two interference types in the behavioral data.

On the EEG side, we analyzed the data both using stimulus-locked and response-locked methods. In both cases, we first conducted a whole-scalp ERP analysis to get a broad picture of the activity over all electrodes and the whole time-window without restricting the analysis to specific components. In line with previous studies, we expected lower amplitudes in similar than dissimilar contexts, for both semantic and phonological manipulations. Our first critical prediction relates to the similarity between the profile of interference induced by contextual similarity and Stroop-like phase manipulation. If the two types of interference are similar in nature, we expect similar electrophysiological profile, e.g., the relative polarity of the condition with greater interference compared to the one with less interference should be similar across contextual similarity and phase manipulation.

For each of the stimulus-locked and response-locked analyses, we also conducted an additional analysis to better compare the mechanisms underlying these two types of interference. For the stimulus-locked analysis, we conducted a Representational Similarity Analysis (RSA; Kriegeskorte et al., 2008). The application of RSA to contextual similarity is straightforward. We expect greater similarity scores in the semantically similar and phonological similar, compared to the unrelated, condition. But there is also evidence from fMRI studies that RSA can distinguish between high- and low-conflict states in tasks such as Stroop. For example, Freund, Bugg, and Braver (2021) showed that in medial frontal regions, high-conflict trials were, on average, more similar to one another than low-conflict trials. They concluded that high-conflict trials shared a common abstract component (conflict) irrespective of the target and distractor features that drove the high similarity scores in RSA. By the same logic, we may expect greater similarity in the reverse, compared to the straight, phase of our Stoop-like task. Additionally, we can make predictions about the change in similarity scores as a function of the application of control. By definition, control is recruited to resolve conflict, i.e., to change a high-conflict situation into a low-conflict one by suppressing

the competitor. In RSA terms, this means a reduction in the discrepancy between the similarity scores of high- vs. low-conflict states, and potentially even flipping similarity if the application of control creates the purest form of a low-conflict trial by completely eliminating the activation of the competitor. Again, if both types of interference are handled similarly by the brain, we would expect similar RSA patterns for contextual similarity and Stroop-like phase manipulations.

Finally, in the response-locked ERPs, we extracted the medio-frontal components known to index performance monitoring. The most famous of these is the error related negativity or the ERN (Gehring et al., 1993; see also Riès et al., 2011, 2013, for ERN in language production) but since our tasks are simple and language production is a highly practiced task, we did not expect many errors. We, therefore, aimed to examine an ERN-counterpart on correct trials, often referred to as the correct response negativity or the CRN (Gehring et al., 2018). There is still much debate about the origin of ERN and CRN and whether they index the same or different underlying processes (see Vidal et al., 2022, for a review). The key finding regarding the CRN are as follows: the latency of the CRN lies between the ERN for partial and complete errors, and its topography is similar to the ERN after Laplacian transformation (Meckler et al., 2017). In neurotypical individuals, the CRN usually has a lower amplitude than the ERN, unless responses are unexpected (Meckler et al., 2011). Similar to the ERN, the CRN is sensitive to performance. For example, the magnitude of the CRN is usually lower on a trial preceding an error trial compared to a correct trial (Allain et al., 2004). Moreover, the amplitude of both the ERN and the CRN grows larger when participants are instructed to monitor their performance more closely (Grützmann et al., 2014).

All these similarities have led to the conclusion that CRN is indeed an index of performance monitoring. However, while the properties of the ERN have been extensively studied, much less is known about the factors modulating the CRN. Likewise, theoretical accounts often focus on explaining the ERN rather than the CRN. Briefly, two general types of models exist: the first type is the conflict model (Botvinick et al., 2001), which links ERN to the amount of conflict between competing representations. This model naturally accommodates CRN, although it predicts its timeline to be earlier than the ERN (Yeung et al., 2004). The second type includes several models all of which are based on reinforcement learning. Three are noteworthy. The original reinforcement learning model (Holroyd & Coles, 2002) considers ERN to be the signature of an outcome that is worse than expected. The

predicted response-outcome model (Alexander & Brown, 2011) is similar in spirit, but instead of linking the ERN uniquely to worse-than-predicted outcomes, considers it to be sensitive to any unexpected outcome, better or worse than predicted. A closely related model, the reward value and prediction model (Silvetti et al., 2011), predicts that both the ERN and the CRN mark unexpected events, hence the usually observed larger magnitude of the former, which is only due to the fact that errors are generally rarer than correct responses.

Our design provides a good test for the conflict account. If the CRN is sensitive to conflict, we would expect a more negative CRN in the reverse compared to the straight phase. Likewise, semantic and phonological similarity should lead to the increased magnitude of the CRN. On the other hand, since the manipulations are not expected to generate many unexpected outcomes, the reinforcement learning models predict that we might only observe a small CRN, insensitive to the manipulations. Here too, similar to the whole-scalp analyses and the RSA, the comparison between the effects of contextual similarity and Stroop-like phase manipulations on the CRN can inform us about the overlap in the underlying processes that handle these two types of interference. If both manipulations induce the same kind of response to increased conflict, we expect similar change to the CRN in reversed vs. straight, as well as in semantic and phonological vs. unrelated conditions. If not, we would expect different effects of the two manipulations on the CRN.

Methods

Participants

Sample size was selected to be similar to EEG studies that manipulated semantic and/or phonological similarity (Feng et al., 2021; Wang et al., 2018). Thirty individuals (16 females) participated in this study. One was excluded because of a history of speech impairment. Participants were all right-handed native English speakers, had normal or corrected-to-normal vision, and were 18 to 35 years old ($M = 24.1 \pm 4.9$). Participants were consented under a protocol approved by the Institutional Review Board of Johns Hopkins School of Medicine and received payment for their participation.

Materials

Ten monosyllabic target words were selected, such that each word (e.g., cake) could be uniquely paired with a semantically related word (e.g., pie), a phonologically related

word (e.g., rake), and an unrelated word (e.g., bear), for a total of 15 unique pairs. This procedure resulted in a fully balanced design, such that each picture appeared in all conditions (see Appendix A for the materials). Semantically related words were coordinates from a semantic category (cake/pie). LSA similarity was computed using an online database (http://lsa.colorado.edu/; Landauer et al., 1998). LSA was higher for the semantic condition (mean = 0.31 ±0.18) than for the unrelated (mean = 0.12 ±0.08) or the phonological (mean = 0.18 ±0.11) conditions. Since Nozari et al. (2016) showed that phonological overlap over non-onset segments, but not over onset segments created interference, phonological overlap was manipulated over non-onset segments of the words (e.g., cake/rake) and ranged from 1 to 2 phonemes in the same position. There was no phonological overlap between pairs in the unrelated and semantic conditions. Ten black-and-white line drawings corresponding to the selected words were chosen from the Snodgrass database or Google images and scaled to 300 x 300 px for presentation to the participants.

Procedures
Participants were seated approximately 25 inches from a 15-by-12 inch Dell monitor. Their verbal answers were recorded with a digital recorder (Sony ICD-PX333) and an Audio-Technica microphone, and their EEG data were simultaneously recorded.

The paradigm (Figure 1) was adapted from Nozari et al. (2016) and presented using Presentation (NeuroBehavioral Systems). The experiment comprised 15 blocks, 5 blocks of each relation (semantic, phonological and unrelated), presented in a pseudo-randomized order. Each block contained pictures from one pair (see Materials). A block had two phases, straight and reverse. In the straight phase, participants first saw the pair of pictures and their respective labels and named them out loud. On the next four practice trials, one of the two pictures was presented on each trial and participants named it as quickly and accurately as they could. This was followed by the experimental phase with 16 presentations of the same pictures (8 of each, in randomized order). This concluded the straight phase of the block. The following reverse phase followed the same steps, except that participants were now instructed to switch the name of the pictures, e.g., to say "pie" when they saw "cake". The same number of practice and experimental trials were completed in the reverse phase. Each picture was displayed for 1.5 seconds and replaced by a fixation cross displayed for an average of 2 seconds (jittered from a uniform distribution between 1.8 and 2.2s), before the next picture appeared.
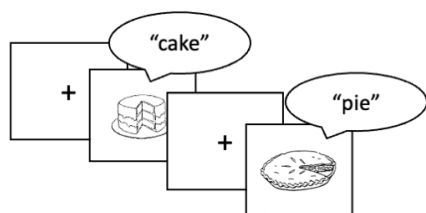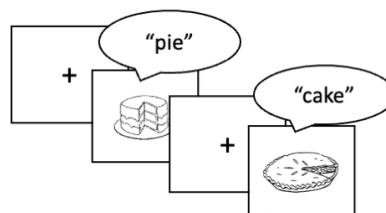
Figure 1. An example trial in the straight and reverse phases of the paradigm for one of the semantically related blocks with cake/pie.

EEG recordings and preprocessing

EEG data were recorded using a 128-channel HydroCel Geodesic Sensor Net (EGI system), with a vertex reference. Data analysis was performed with Brainstorm (Tadel et al., 2011). Data were filtered offline (0.1-100 Hz) and re-referenced to the average of all channels. Noisy channels were rejected and ocular artefacts were corrected with ICA Infomax (Makeig et al., 1995). Manual artefact rejection was performed before segmentation. Data were segmented in stimulus-locked epochs from -500 to 1000ms and response-locked epochs from -1000ms to 500ms (locked on vocal onset, see below). Blind Source Separation based on Canonical Correlation Analysis (BSS-CCA) was run after segmentation (De Clercq et al., 2006; De Vos et al., 2010; Riès et al., 2021). A baseline by subtraction was applied from -200ms to 0 for stimulus-locked epochs, and the same pre-stimulus baseline was used for response-locked epochs.

Analysis

Behavioral data

Responses were transcribed by a native English speaker naïve to the hypotheses of the study and checked by a second coder. RTs (calculated as the time from stimulus presentation to vocal onset) and spoken durations were extracted using Praat (Boersma & Van Heuven, 2001), and manually double-checked. RTs and durations were analyzed only on correct trials and values outside 3SD from the individuals' means were rejected. Following a Box-Cox test, both RTs and durations were log-transformed for the analyses. Data analysis was performed using mixed linear models (Baayen et al., 2008) in R version 4.0.5 with package lmerTest (Kuznetsova et al., 2017). We aimed for a maximal random effect structure (Barr et al., 2013), but due to convergence issues, and to keep models comparable, we only included random

intercepts of subjects and items in all models. We used treatment contrasts without centering, with the unrelated and straight conditions as reference levels, respectively. Each contrast was also tested with the Wilcoxon non-parametric test, to double check the results of the mixed models.

EEG data

EEG analysis was performed on correct trials only. Trials were excluded if they contained mumbling, filled pauses (uh/um), or additional utterances (e.g., the) before the correct response was given, to prevent premature articulatory activity from contaminating the signal of interest. We performed three sets of analyses on the EEG data. Each set was conducted once for the Stroop-like phase manipulation and once for the contextual similarity manipulation. The former analysis compared the EEG signal in straight vs. reverse phase. The latter tested two contrasts, semantic vs. unrelated, and phonological vs. unrelated. Note that semantic and phonological contexts are not directly comparable, as they differ in more than one way. The three sets of analyses are described below.

(1) A whole-scalp analysis, conducted once by time-locking the EEG data to the onset of the stimulus (stimulus-locked analysis) and once to the onset of articulation (response-locked analysis), without assuming a particular topography or timeline a given effect. To test statistical significance, we performed non-parametric point-by-point comparisons using the Wilcoxon signed-rank test. P-values were calculated by permutations (N = 1000) in the Brainstorm software (Tadel et al., 2011). Time-windows of significant results are reported at $p<.05$.

(2) An RSA analysis computed using custom-made scripts. By-item ERPs were computed in each condition for each participant (average of 8 presentations). Then each item ERP was correlated with the item ERP of the same pair (for all 15 pairs present in the experiment) at each time point in each condition, for each participant. The resulting arrays were averaged across pairs by participant, and subsequently across participants to derive the average similarity score for each condition. The same procedure was followed for phase reversal, averaging over 45 pairs (15 pairs in each of the three context conditions that make up the straight and reverse condition respectively). Statistical analysis was conducted using the Wilcoxon signed rank test and permutations.

(3) A CRN analysis, along with other medio-frontal components was conducted by a Laplacian transform using spherical splines, order 4, degree 20, lambda = 10^{-5}. Applying a Laplacian transformation is desirable, because it acts as a high-pass spatial filter, which reduces volume conduction effects and, consequently, overlapping effects in both time (Burle et al., 2015) and space (Nuñez & Srinivasan, 2006). In fact, Laplacian transformation has been instrumental in uncovering the CRN (Vidal et al., 2000). Statistical analysis was conducted over FCz electrode, using Wilcoxon signed rank test and permutations.

All analysis scripts are available here: https://osf.io/vmskx/.

Results

Behavioral results

We collected 13,898 responses, of which 13,663 were correct. Outlier rejection resulted in the discarding of 1.3% of data. There were significantly more errors in the reverse compared to the straight phase, β = -1.53, z = -5.44, p <.001 (Figure 2, left panel). The main effect of semantic similarity on errors was marginal (β = -0.56, z = -1.75, p =.081). The effect of phonological similarity did not reach significance. Neither did the interaction between phase and similarity (see Table B2 in Appendix B for the full results of this analysis).

RTs were significantly longer in the reverse (700 ± 94ms) than the straight (603 ± 79ms) phase (β = 0.14, t = 25.1, p <.001), but there were no significant effects of semantic or phonological similarity on RTs, or any interactions between similarity and phase (Figure 2, middle panel; see Table B4 in Appendix B for the full results of this analysis). Spoken durations were marginally longer in the straight vs. the reverse phase (β = -0.01, t = -1.96, p = .050). There were also significant main effects of semantic (β = 0.02, t = 3.08, p = .002) and phonological similarity (β = 0.014, t = 2.94, p =.003), with both resulting in longer durations compared to the unrelated baseline (Figure 2, right panel). The interaction of phase and phonological similarity was also significant on durations (β = 0.02, t = 2.82, p =.005; see Table B6 in Appendix B for the full results). Post-hoc tests, unpacking this interaction, revealed that there was a significant effect of phonological similarity in both the straight (β = 0.02, t = 3.74, p < .001) and in the reverse (β = 0.03, t = 6.25, p < .001) phases. The interaction thus points to the larger magnitude of the phonological interference in the reverse than in

the straight phase, rather than the absence of the effect in one phase.

Summary of the behavioral analyses. As expected, the Stroop-like effect of phase on behavioral data was strong and apparent on errors, RTs, and (marginally) on durations. The effect of similarity was more subtle but was observed robustly on durations for both semantic and phonological similarity, and also marginally on accuracy for semantic similarity. The effects of similarity were observed in both straight and reversed phases, with phonological similarity showing a greater effect in the reversed phase. There were no other statistically significant interactions. All the statistically significant effects were confirmed by non-parametric tests. These, along with full tables of the MLM analyses, can be found in Appendix B.
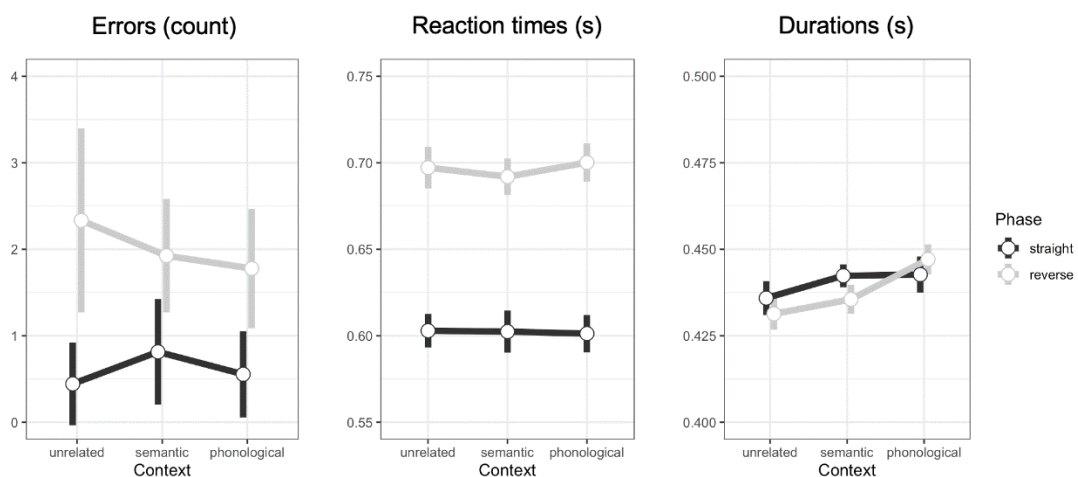

Figure 2. Means and SEs of error counts, reaction times, and durations by similarity (unrelated, semantic, phonological) and phase (straight, reverse).

EEG results

Stimulus-locked analyses

Whole-scalp analyses. The effects of phase reversal were widespread, much larger than the effects of contextual similarity, and comparable for different similarity conditions. Therefore, we present the results with all similarity conditions averaged. Effects were evident over the whole time-window: Figure 3A displays the t-values (thresholded for $p < .05$) by channels over time. The subplot below shows the number of electrodes (out of 128) displaying a significant effect. Both figures point to a widespread effect that can be described in two stages. From 200ms to 350ms, higher amplitude was observed for reverse than straight naming over central electrodes. The difference in amplitude between reverse and straight conditions on Cz electrode (Figure 3B) was significantly correlated with the interference effect (reverse - straight)

observed on reaction times, $r = -.48$, $p = .010$, but not on duration, $p > .05$. After 400ms, two opposite effects were observed and persisted over the whole time-window: a negative effect over central and posterior electrodes and a positive effect frontally. Importantly, the nature of the components was similar in both the reverse and straight phase. The main observable changes were in terms of latency, suggesting that even in the reverse phase, the same processes underlying word retrieval were likely taking place.

To test the effects of contextual similarity, we focused the analyses only on the straight phase for three reasons: (1) Focusing on the straight phase allows for a cleaner comparison between contextual similarity and Stroop-like effects. (2) The effect of reversal is so overpowering that it obscures effects of contextual similarity in the reverse phase. (3) Finally, focusing on the straight phase allows us to compare the effect to previous reports. For the semantic context, we observed a significant effect over parietal electrodes from 250ms to 350ms, in a time-window similar to the N2 described in previous work (Figure 4A). The unrelated context was significantly more negative than the semantic context. A later effect, this time showing a centro-parietal topography was observed from 550ms onwards. In that time-window, again, the semantic context had a significantly lower amplitude than the unrelated context visible on more than 20 electrodes (see Figure 4B for examples shown on Cz and Pz electrodes). For the phonological context, we observed a fronto-central effect around 300 to 350ms with a lower amplitude for the phonological than the unrelated context over more than 10 electrodes (Figure 4D; see also Figure 4C for examples shown on Cz and FCz electrodes). No later effect could be detected. After correction for multiple comparisons, the correlation between EEG and behavioral data did not reach significance for contextual similarity.

RSA. The RSA for phase revealed a pattern in two phases (Figure 3C). In a first phase, the reverse condition displayed significantly higher similarity than the straight condition from 250 to 450ms (point-by-point Wilcoxon rank test thresholded at $p < .05$). In a second phase, the pattern reversed with the reverse condition exhibiting significantly *lower* similarity than the straight condition for the rest of the time-window of interest (see Figure 3C for the window demonstrating significant differences). The RSA for contextual similarity showed that both semantic or phonological similarity also induced significantly higher similarity scores starting from 100ms and across the whole time-window, compared to the unrelated condition (Figure 4E). However, no reversal of the

pattern was observed later on, with both semantic and phonological contexts keeping a significantly higher similarity score than the unrelated context up to articulation.
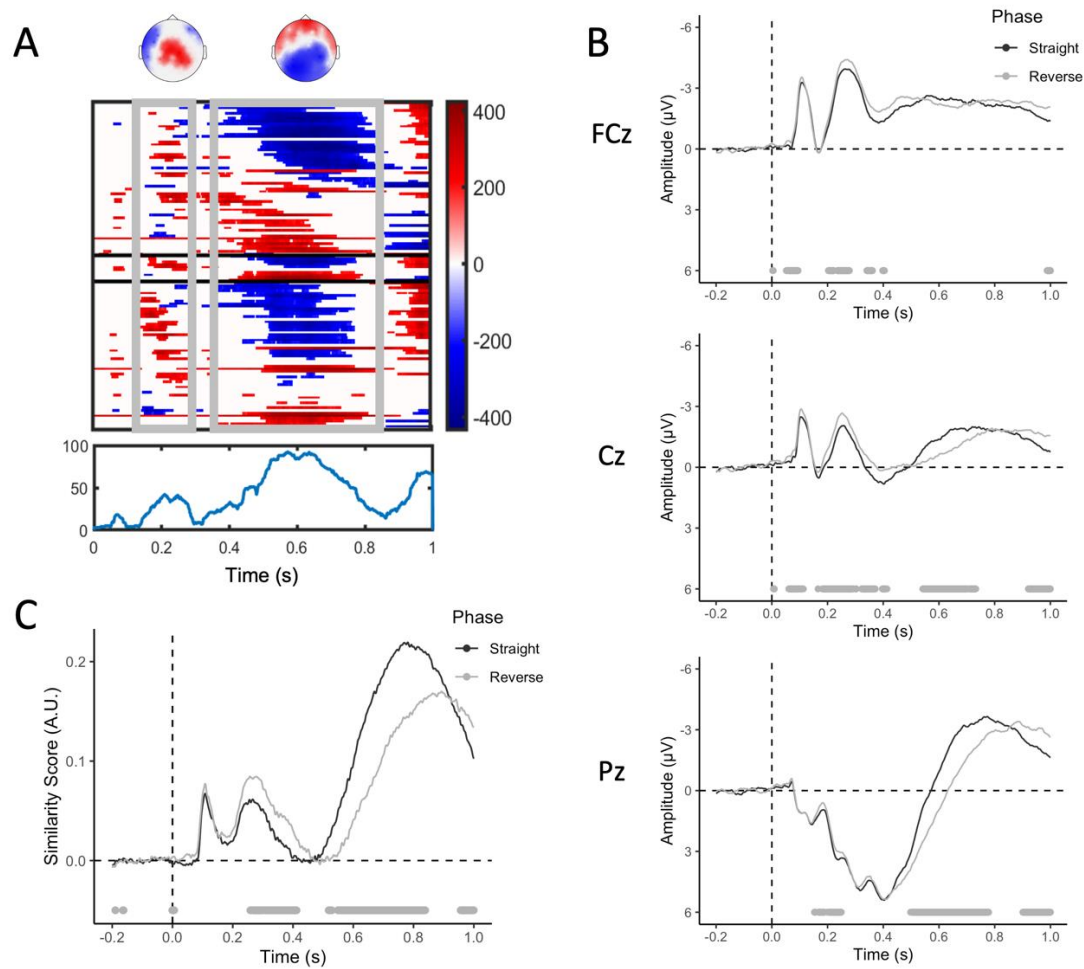

Figure 3. Stimulus-locked ERP results for straight and reverse naming. (A) Channels-by-time map of the significant statistical effects (t-values, thresholded for *p* < .05). Electrodes (y-axis) are ordered from posterior to anterior within left hemisphere, midline and right hemisphere. The effects of interest (see main text) are outlined in grey frames. The number of electrodes with a significant effect is plotted over time below. (B) ERPs averages for midline electrodes (FCz, Cz, Pz). (C) Results of the Response Similarity Analysis (RSA): similarity score for straight and reverse phase. Gray dots/bands at the bottom of the plot correspond to significant Wilcoxon rank tests at the .05 level.
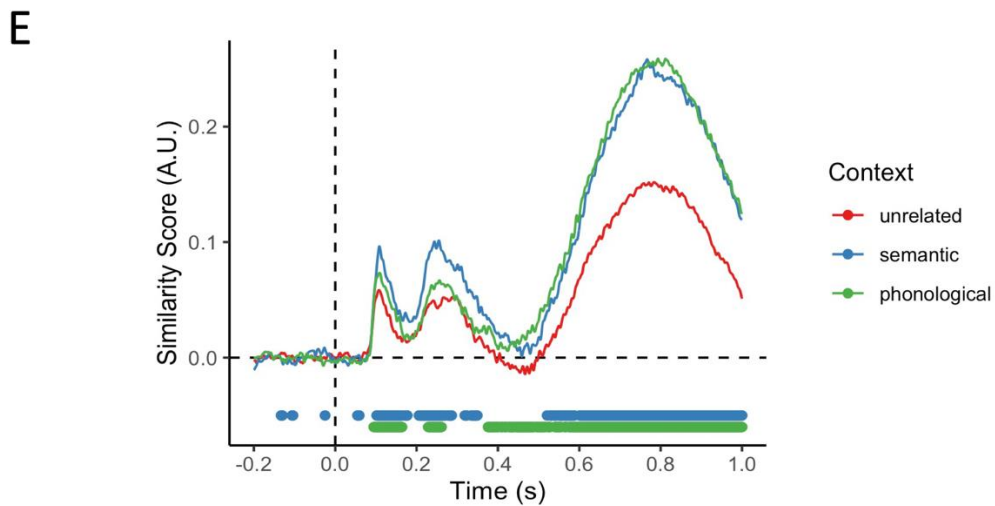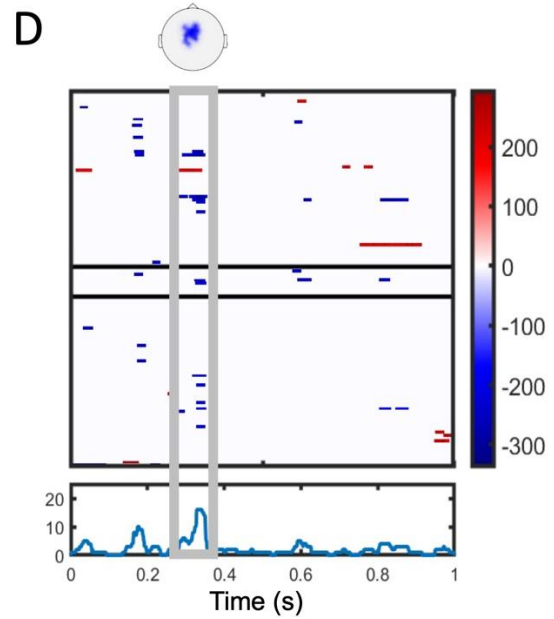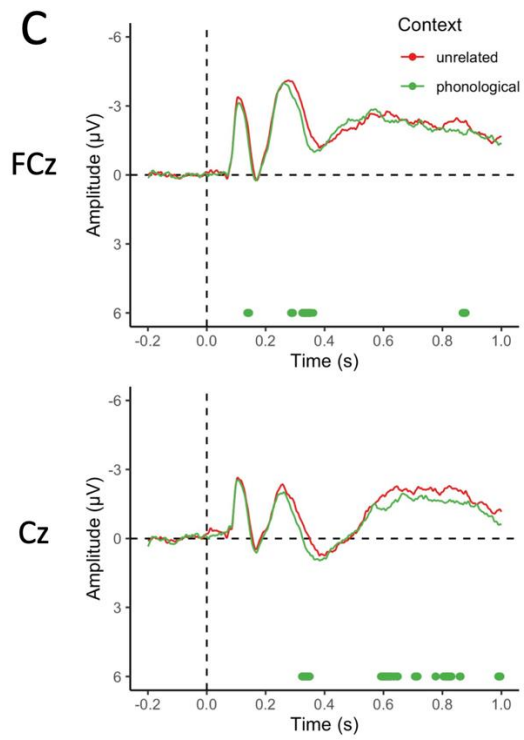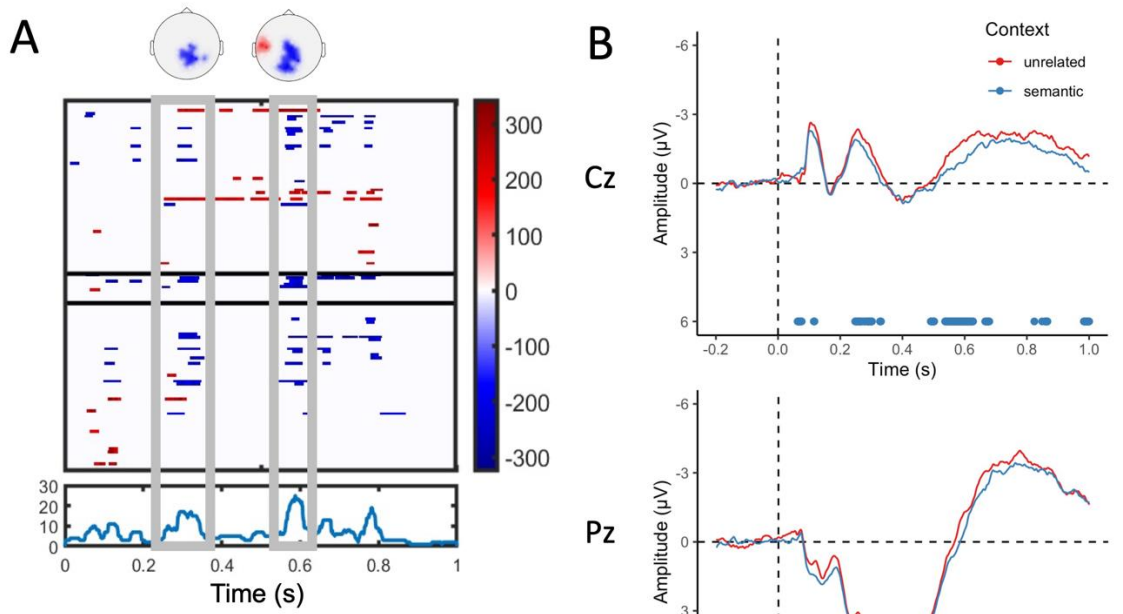
Figure 4. Stimulus-locked ERP results for semantic (A, B) and phonological (C, D) contexts. Each related context (semantic in blue, phonological in green) is compared to the unrelated (red) context. A, D: Channels by time map of the significant statistical effects (t-values, thresholded for *p* < .05). Electrodes (y-axis) are ordered from posterior to anterior within left hemisphere, midline and right hemisphere. The effects of interest (see main text) are outlined in grey frames. The number of electrodes presenting a significant effect is plotted over time below. B, C: Midline electrodes (FCz, Cz, Pz) displaying the effects of interest. (E) Results of the Response Similarity Analysis (RSA): similarity score for unrelated (red), semantic (blue) and phonological (green) context. Dots and bands at the bottom of the plot correspond to significant Wilcoxon rank tests at the .05 level, for the contrast of semantic vs. unrelated (blue) and phonological vs. unrelated (green).

Summary of the stimulus-locked analyses. We observed the effects of both phase and similarity manipulations on the EEG signal locked to the stimulus onset. Stroop-like phase reversal showed widespread effects over the whole time-window and scalp, whereas contextual similarity effects were more local. While the effects of semantic and phonological similarity overlapped in time, the former started earlier and had a different topography than the latter. Importantly, RSA results showed initial similarities between phase and contextual similarity effects: the reverse phase, as well as semantically and phonologically related blocks, showed greater similarity in the initial stages of processing, indexing greater overlap between the activation of the two competing items. Most interesting, however, was the difference observed between the RSA pattern for phase and contextual similarity manipulations: phase manipulation showed a biphasic pattern, in which the initially greater similarity of the reverse phase switched to less similarity for items in the reversed block. This is compatible with a signature of control, which actively suppresses the competitor. No such switch was observed in the semantically or phonologically related blocks, with higher similarity compared to the unrelated condition observed throughout processing.

While we expected that the application of control should decrease the discrepancy between the similarity scores on straight and reverse trials, a complete reversal of similarity scores across the straight and reverse phases may be less expected. There are two potential explanations for this prominent reversal: First, target activations are never pure. There is always some degree of "noise" in the system, which activates non-target representations. It is possible that the application of inhibitory control in the high-conflict condition reduces the noise to a near-zero level by strongly suppressing

competing alternatives. If so, the end-state of a high-conflict trial after the application of control can be even cleaner with respect to conflict than a low-conflict trial which did not enjoy the benefit of control-driven clean-up of noise. This would manifest as lower similarity scores in the reverse phase around articulation time. A second explanation is that participants benefit from the coactivation of both items within the block in working memory, as participants know that they will reoccur. Therefore, the simultaneous activation of the competitor may not be only due to noise, but to some extent by strategy. Therefore, when control strongly suppresses the competitor, the now-low-conflict reverse condition is compared to a straight condition that maintains the activation of the competitor to some extent to use it for future trials. Again, this would manifest as lower similarity scores in the reverse phase in later stages of the task. The important point is that we do observe the reduction of the discrepancies in RSA scores between high and low-conflict conditions over time in phase reversal, compatible with the active application of inhibitory control, where no such pattern is observed in contextual similarity. If anything, the magnitude of the difference increases over time between phonological and unrelated conditions (Figure 4E).

In short, both analyses revealed effects that extended over 600 ms post-stimulus onset, which points to the availability of information across the full window of word production. However, the RSA showed very different patterns for Stroop-like phase and contextual similarity manipulations, suggesting that the system does not handle the two kinds of interference in the same way. In response-locked analyses, we again provide a comparison of the timeline of phase and similarity manipulations, but also focus on a specific component of general monitoring performance, the CRN.

Response-locked analyses
Whole-scalp analysis. On response-locked ERPs, the effect of Stroop-like phase reversal was again widespread and evident over the whole time-window (Figure 5A, see Figure 5B for example electrodes). From -900ms to -400ms, higher amplitude was observed for reverse than straight naming. The spatial distribution of this effect was similar to that of the late effect observed in stimulus-locked averages: negative over posterior electrodes, and positive over frontal electrodes. Closer to vocal onset, from -300ms to 0 another peak in the number of significant electrodes was observed, and polarity was reversed with the straight condition showing higher amplitude than the reverse condition. This effect (difference in amplitude between reverse and straight) was again significantly correlated with the interference effect observed on reaction

times, r = -.59, p < .001. Only sparse differences were observed after vocal onset.

A similar analysis was conducted to test the effects of contextual similarity. In the semantic context, significant differences were observed over the whole time-window, first, and more pronouncedly, over parietal electrodes from -500ms to -300ms, and later over central electrodes from -100ms to 100ms (Figure 6A, 6B). Overall, the semantic context showed less negative amplitude than the unrelated context. In the phonological context, the most prominent effect was observed around the response from -100 to 150ms over centro-parietal electrodes (Figure 6C, 6D). Another more scattered effect was observed earlier from -550 to -350ms over left parietal electrodes. The direction of the effect was the same as semantic context. After correction for multiple comparisons, the correlation between amplitude on central electrodes and behavioral indices did not reach significance.

CRN. The task did not generate enough errors for a separate comparison of ERN and CRN. After Laplacian transform, we observed a clear CRN component for both the straight and reverse conditions, peaking right after speech onset, but the amplitude of the CRN was not sensitive to phase (Figure 5C). However, prior to the CRN, the reverse condition displayed a significantly higher amplitude than the straight condition, from -500ms to -150ms, over a component that has been described previously in language production (Riès et al., 2013) and has exactly the same topography and timeline as the one we report here, peaking at -300ms before the response. Between -350 and -150ms, the difference in the amplitude of the reverse and straight phases showed a significant negative correlation with RT, r = -0.60, p < .001, but not with duration.

Similar to phase, semantic similarity did not influence the CRN amplitude. In contrast, phonological similarity induced a significantly *lower* amplitude than the unrelated condition after vocal onset from 0 to 100ms (Figure 6E; point-by-point statistical analysis thresholded at *p*<.05). Unlike phase, contextual similarity did not lead to any differences between conditions prior to the emergence of the CRN.
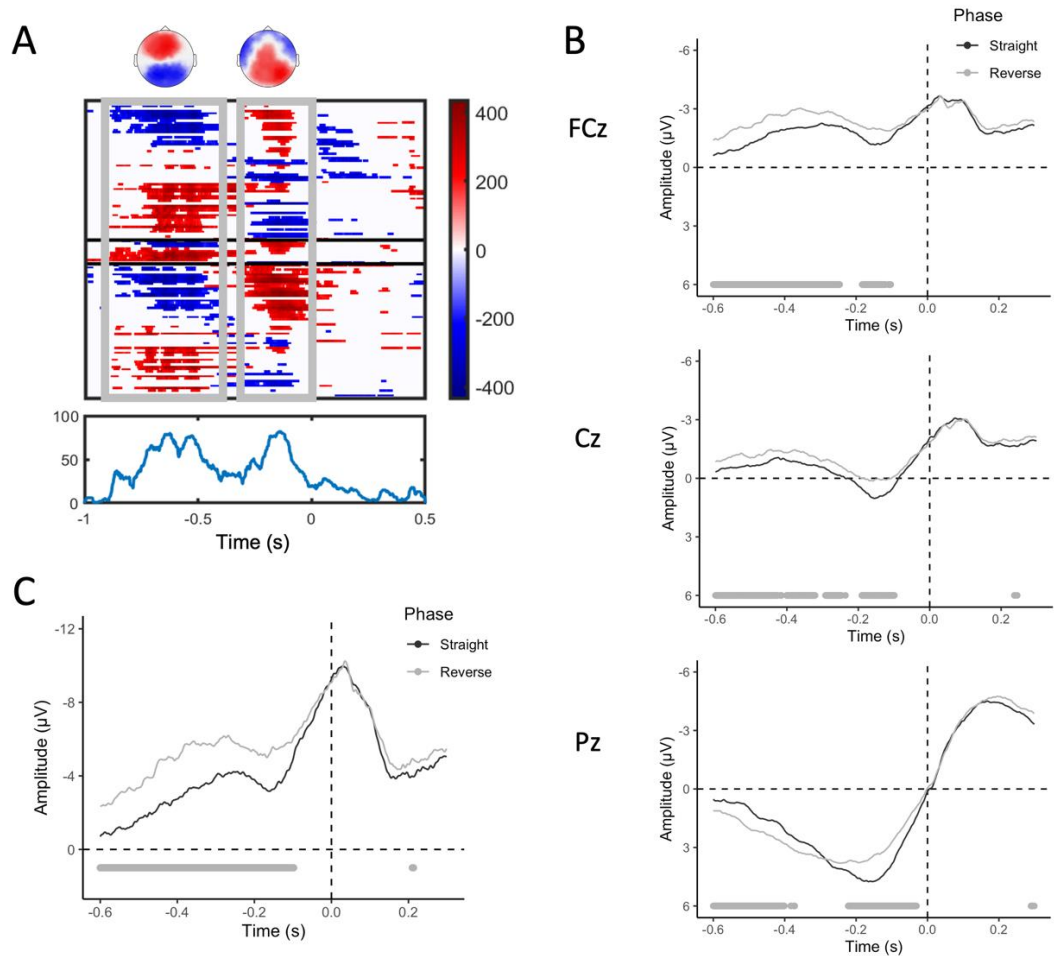
Figure 5. Response-locked ERP results for straight and reverse naming. (A) Channels by time map of the significant statistical effects. Electrodes (y-axis) are ordered from posterior to anterior within left hemisphere, midline and right hemisphere. The effects of interest (see main text) are outlined in grey frames. The number of electrodes presenting a significant effect is plotted over time below. (B) Midline electrodes (FCz, Cz, Pz). (C) Response-locked Laplacian-transform for phase reversal at FCz.

23

Figure 6. Response-locked ERP results for semantic (A,B) and phonological (C,D) conditions. Each condition (semantic in blue, phonological in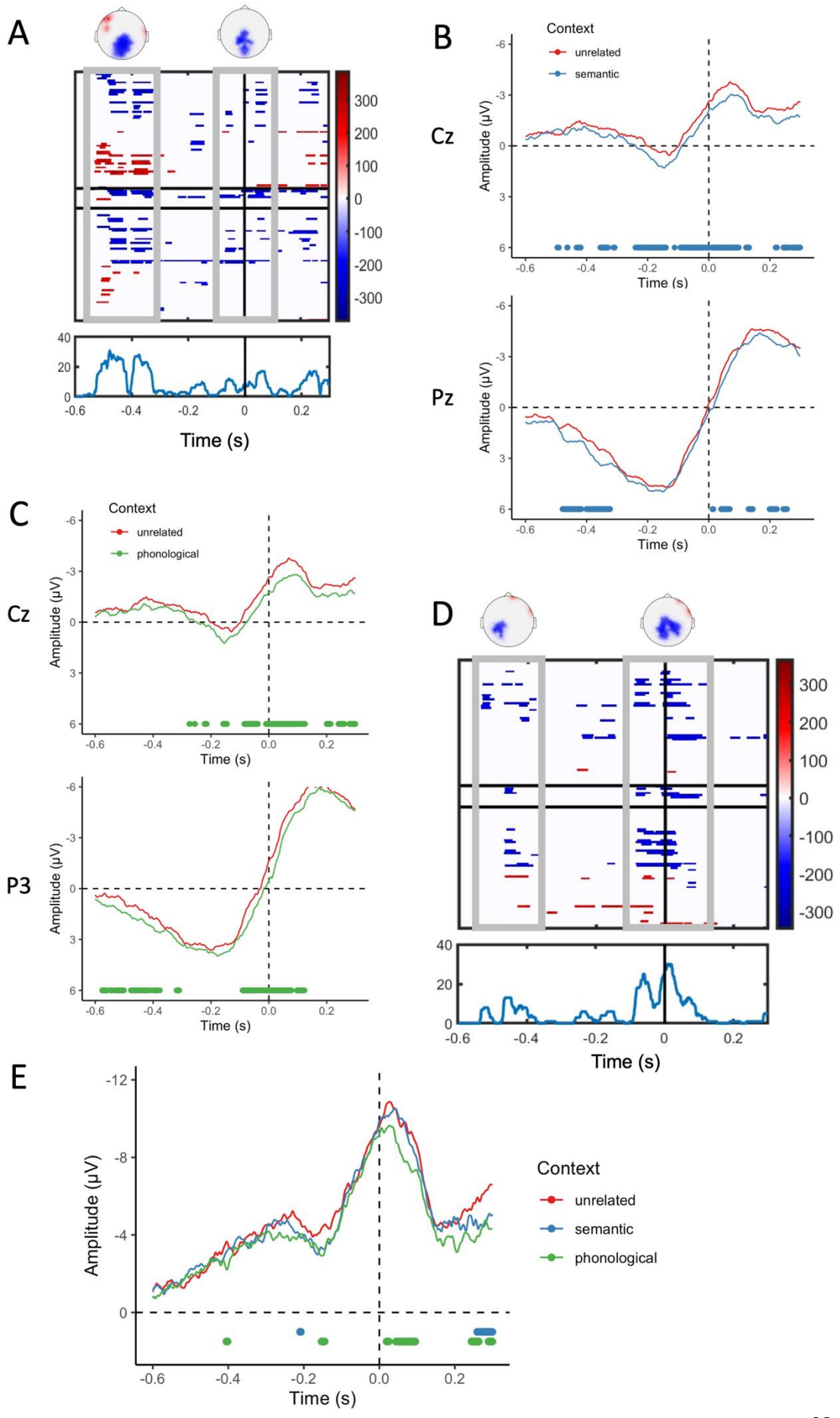 green) is compared to the unrelated (red) condition. (A,D): Channels by time map of the significant statistical effects. Electrodes (y-axis) are ordered from posterior to anterior within left hemisphere, midline and right hemisphere. The effects of interest (see main text) are outlined in grey frames. The number of electrodes presenting a significant effect is plotted over time below. (B,C): Midline electrodes (FCz, Cz, Pz) displaying the effects of interest. (E) Response-locked Laplacian-transform activity at FCz for context manipulations.

Summary of the response-locked analyses. Similar to what we observed in stimulus-locked analyses, the effects of both phase and contextual similarity were observed over a large window, suggesting that information about differences between conditions is available throughout the processing of a word. Moreover, the effects of semantic and phonological similarity overlapped substantially in time. However, and also similar to the pattern observed in stimulus-locked analyses, semantic context had a greater impact in the early stages and phonological context towards the later stages of processing.

Analysis of medio-frontal components revealed a clear CRN component in all conditions. Critically, the amplitude of CRN was neither sensitive to Stroop-like phase reversal, nor to semantic similarity. In contrast, we observed a significantly lower-amplitude CRN for the phonological condition compared to the unrelated condition. Moreover, we found clear differences on an earlier component for the Stroop-like phase reversal, but not for contextual similarity manipulation.

Discussion

The study investigated the electrophysiological signatures of two types of interference in language production, a Stroop-like interference imposed by reversing the names of two pictures within a block, and a similarity-induced interference by increasing competition from a semantic or a phonological competitor (Mirman & Magnuson, 2008; Sadat et al., 2014). We deliberately used a paradigm that accommodated both manipulations within the same task and items, without resorting to processes that are not a routine part of language production in conversations, such as word reading (e.g., as in picture-word interference). As expected, behavioral interference was stronger for the Stroop-like phase reversal and observed in error rates, RTs, and durations. A more subtle but robust interference effect was observed on durations for both semantically

and phonologically similar contexts. With the exception of an interaction between phase and phonological similarity on word durations (greater phonological interference in the reverse phase), the two manipulations yielded independent effects, in line with the findings of Nozari et al. (2016).

The novel contribution of this study was the comparison of the electrophysiological signatures of these two types of interference. Whole-scalp EEG analyses showed widespread effects in time and space for both Stroop-like phase and contextual similarity manipulations, in both stimulus-locked and response-locked analyses. These effects, however, were more localized for the similarity manipulations and although overlapping in time, showed a difference in timeline with greater early effects of semantic and greater later effects of phonological similarity. Together, these findings are compatible with a globally modular and locally interactive language production system. We elaborate on this claim in the section "Implications for modularity vs. interactivity".

Two specific sets of analyses were performed to further investigate the similarities and differences between the effects of phase and similarity manipulations, a stimulus-locked RSA and a response-locked CRN analysis. The RSA results showed an initially similar pattern between the effect of Stroop-like reversal and increased competition by similarity. In both cases, greater similarity was indexed for the more interfering conditions, showing increased representational overlap between the competing alternatives in the brain. Most interesting, however, was the difference observed between the results of the RSA for phase and similarity. Stroop-like reversal showed a clear biphasic pattern, with greater similarity for the reversed phase switching to less similarity at around 450-500 ms post-stimulus-onset. No such flip was observed either for the semantically similar or for the phonologically similar conditions. The initial similarity between the patterns observed in the RSA of the two manipulations, followed by clear differences in handling competition in these two cases strongly questions the assumption that both manipulations tap into similar control processes. We further discuss the implications of these results under "Implications for mechanisms of interference and control in language production".

Finally, all conditions generated a CRN. The magnitude of the CRN was, however, not sensitive to Stroop-like phase reversal or semantic similarity. The only manipulation that affected CRN's amplitude was phonological similarity, and it actually *reduced* it. We will unpack this effect under "Implications for CRN and general monitoring".

Implications for modularity vs. interactivity

The effects of contextual similarity have been previously taken to argue for serial vs. interactive models of language production. The effects observed in the current study bridge the gap between multiple seemingly divergent results of the past studies. First, the timeline of the semantic similarity effect seen in the current study is compatible with past reports of an early effect (e.g., Costa et al., 2009), as well as a later effect (e.g., Janssen et al., 2015). The direction of the effect (lower amplitudes for related than unrelated) is also consistent with the existing literature (Janssen et al., 2011; Python et al., 2018a). Second, the stronger locus of the semantic vs. phonological effect on earlier vs. later stages of production is compatible with Wang et al. (2018), even though they manipulated phonological onset and thus observed behavioral facilitation for phonological overlap. Third, the window of overlap observed between semantic and phonological effect is compatible with a recent study by Feng et al. (2021).

The finding of overlap between semantic and phonological manipulations is compatible with a system that is not entirely serial and modular (e.g., Feng et al., 2021; Riès et al., 2017). The late effects of semantic similarity show cascading in the production system, i.e., continuing activation of semantically activated lexical items during phonological encoding and possibly even articulatory phonetic encoding. The earlier influence of phonological similarity is compatible with a system that rapid spread of information in all layers of the system (Miozzo et al., 2015; Strijkers & Costa, 2016) and/or feedback from the phonological layer affecting lexical selection. At the same time, the greater influence of semantic similarity on earlier stages of production vs. the more pronounced effect of phonological similarity on later stages of production (especially clear in the whole-scalp response-locked analyses), together with differences in their topographies, speaks against a fully interactive system with no modularity. Rather this aspect of the data is closer to the classic findings of early semantic processing followed by a later stage of phonological encoding (e.g., Indefrey, 2011; Indefrey & Levelt, 2004).

Collectively, these results point to a system that is globally modular system, with semantic-lexical processing generally preceding phonological encoding, but with some degree of cascading and interactivity between the layers, which makes the processing of information at different layers overlap in time. Such overlapping activity allows for some influence of information from lexical selection on phonological encoding and vice

versa, as have been suggested by behavioral data (Dell, 1986; Nozari & Dell, 2009; Rapp & Goldrick, 2000; see Dell et al., 2014, for a review of the behavioral evidence).

Implications for mechanisms of interference and control in language production

As shown in many previous studies, the current results are compatible with increased production difficulty as a function of both semantic and (non-onset) phonological overlap (Belke et al., 2005; Breining et al., 2016; Feng et al., 2021; Nozari et al., 2016; Qu et al., 2021; Schnur et al., 2006, 2009). Due to this behavioral difficulty and the system's ability to overcome it, it has sometimes been assumed that inhibitory control is involved in resolving conflict between similar representations and that this inhibitory control must be the same as that involved in tasks such as Stroop. Several aspects of the current results are incompatible with this assumption.

First, there was a strong *negative* correlation between the ERP amplitude and RTs for phase manipulation in both stimulus- and response-locked whole-scalp analyses, such that the greater difference in EEG amplitudes of reverse over straight phases, the smaller the Stroop-like cost. This correlation shows that what is indexed by the EEG differences between reverse and straight phases is the implementation of control, which is driving down the behavioral cost of reversal. No such correlation was observed between behavioral and EEG data for contextual similarity. The second aspect is that the direction of the effects of contextual similarity and Stroop-like phase reversal are the opposite of one another in EEG data. Whereas both semantic and phonological similarity generate *lower-amplitude* EEG waveforms compared to the easier unrelated condition (Janssen et al., 2011; Python et al., 2018a), reverse phase generates *higher-amplitude* waveforms on similar electrodes. If one accepts, as strongly implied by the negative correlation between EEG and behavioral data discussed above, that the phase manipulation indexes cognitive control, one expects contextual similarity to induce the same polarity differences, the opposite of what is observed.

The third and most convincing aspect of the data that points to fundamental differences between Stroop-like and similarity-induced effects are the results of the RSA. While both similarity and phase manipulations initially produced higher similarity scores for conditions that created more interference, a clear reversal was observed only for the Stroop-like effect, such that after 500ms, the straight phase was the one with higher

similarity. Similar to the finding of negative correlation between behavioral and EEG data, this flip in similarity scores is a hallmark of the implementation of control. When two items are highly coactivated and interfere with one another in initial stages of the reversal phase, the recruitment of control suppresses the irrelevant response thus reducing similarity in later stages of processing. No such flip was observed in either semantic or phonological conditions. Instead, a higher similarity score was indexed in both conditions compared to the unrelated condition up to articulation.

Collectively, these results show that, similar to the classic Stroop task, the current Stroop-like reversal manipulation recruits control and such control is indexed in the EEG signal and its relation to the behavioral data. However, the electrophysiological signature of contextual similarity is entirely different. It is in the opposite direction, does not show a robust correlation with behavioral data, and does not exhibit the reversal which marks the implementation of control in resolving the Stroop-like interference. While the lack of a correlation could be due to noise, the other two differences cannot be explained by noise or lack of statistical power. Together, these findings imply that the interference observed as a consequence of contextual similarity is not handled by the system in the same way as Stroop-like interference.

Although in the introduction, we justified the choice of the Stroop-like paradigm as a good proxy for the experimental investigation of control in everyday language production, it is still reasonable to wonder if the effect of phase is due to working memory rather than inhibitory control demands. First, note that even in the verbal version of the classic Stroop task, the congruent and incongruent conditions differ in memory demands. Although the ink color is a cue in the incongruent condition, the word must still be retrieved from memory, whereas in the congruent condition, it is right there. Thus the current task is not radically different from the verbal version of Stroop, PWI, or similar paradigms. Rather, the difference is a matter of degree to which memory retrieval is demanding in the incongruent or high-conflict condition. We tried to minimize this burden by having the straight phase (16 trials) precede the reversed phase. There were also 2 introductory trials and 4 practice trials in each phase. This means that by the time that participants had to reverse the two names within a block, they had practiced the pair 28 times. This is ample practice for remembering two words (much more than in most button-press cognitive control tasks in which participants have to commit an arbitrary mapping between buttons and answers to memory).

But let us assume that some difficulty due to the burden of working memory remains. The question is: can it provide an alternative account of the data? The key analysis here is the RSA. Recall that the a priori predictions about the differences in similarity scores between high and low-conflict conditions and the effect of control were derived from a study using the classic button-press Stroop task that did not have the memory demands of the current Stroop-like paradigm. These predictions were borne out in the current data. It is undeniable that our task requires inhibitory control, just like the classic Stroop task. Whether there is an additional influence of working memory or not, the observed pattern is fully predicted by, and compatible with, difference in control demands. Importantly, the same predictions were made for contextual similarity, again purely based on potential differences in control, and the data did not support them, showing that contextual similarity differs even from a classic button-press Stroop task that does not have the memory demands of the current Stroop task. Therefore, although we acknowledge that there may be additional effects of memory load above and beyond cognitive control in our Stroop-like task, these effects do not by themselves explain the differences observed in RSA between the two tasks.

At the behavioral level, the conclusion above implies that robust correlations between interference effects resulting from similarity and Stroop-like manipulations should not be expected. This explains the weak and mixed results of the studies discussed in the introduction that have attempted to demonstrate such correlations. At the theoretical level, the results argue against similar underlying mechanisms for the handling of Stroop-like and similarity-induced effects, with only the former showing a clear signature of control. In fact, the current data are most compatible with models that do not pose an active control mechanism for the resolution of similarity-induced interference, such as incremental learning models of similarity-induced interference (Breining et al., 2019; Oppenheim et al., 2010; Qu et al., 2021). Importantly, an advantage of these models is that they can easily handle other effects of similarity, such as the facilitation induced by thematic relations. For example, Oppenheim and Nozari (2021) simulated the opposite effects of taxonomic and thematic semantic similarity on naming the same target picture in a model with incremental learning, and specifically showed that these opposite behavioral effects were unrelated to a competitive or non-competitive selection rule and active implementation of control.

Implications for CRN and general monitoring
Unlike some of the past studies, our goal here was not to compare ERN and CRN, but

rather to compare CRN across conditions with different degrees of conflict and representational overlap. Compatible with the predictions of the conflict model, we did observe a prominent CRN in all conditions, potentially because two responses are kept simultaneously activated in all blocks of this study. However, contrary to the predictions of the conflict account, CRN's magnitude was not systematically modulated by the amount of conflict. Most prominently, CRN was completely insensitive to the Stroop-like phase manipulation. A possible explanation is the timeline of the effect: as simulated by Yeung et al. (2004), a conflict monitoring account would predict an earlier signature of conflict on correct than error trials. In line with this prediction, we did observed differences in the time window leading up to articulation. A medio-frontal component (N-300) earlier than the CRN had a larger amplitude for reverse than straight naming. Previous reports suggested that this component could be analogous to the N-40 described in choice reaction time tasks that indexes response selection (Riès et al., 2013). But this explanation is insufficient to justify the finding of a prominent CRN at the time of articulation.

The conflict account is also incompatible with the only factor that did modulate the CRN: we found that phonological similarity *decreased* the amplitude of CRN. Recall that the effect of phonological overlap, similar to semantic overlap and phase reversal is behavioral interference. Thus, the lower amplitude of CRN in the phonological compared to the unrelated condition is against the predictions of the conflict account. Instead, the only difference between the phonological and other conditions is the number of articulatory units to be used across the two responses within the block, which is similar in all other blocks and always higher than the phonological block. Since most tasks that are used for studying the ERN/CRN use button press, it is difficult to evaluate the hypothesis that CRN's amplitude may index the number of active planning units for potential responses using the current literature, but our results provide a strong basis for testing this hypothesis in future studies.

Conclusion

Converging evidence showed that the interference resulting from Stroop-like manipulations and contextual similarity do not have similar electrophysiological signatures in the language production system. Importantly, while the former shows evidence of recruiting active control to suppress the competing response, the latter does not. Instead, all electrophysiological differences between similar and unrelated contexts, including the CRN, appear to index the degree of representational overlap,

rather inhibitory control. These findings speak against the models of contextual interference in language production which require active inhibitory control for suppressing the competitors.

Acknowledgments

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*(3), 604–614. https://doi.org/10.1037/0278-7393.33.3.604

Alario, F.-X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 741–764. http://dx.doi.org/10.1080/027249800410535

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience*, *14*(10), 1338–1344. http://dx.doi.org/10.1038/nn.2921

Allain, S., Carbonnell, L., Falkenstein, M., Burle, B., & Vidal, F. (2004). The modulation of the Ne-like wave on correct responses foreshadows errors. *Neuroscience Letters*, *372*(1–2), 161–166. https://doi.org/10.1016/j.neulet.2004.09.036

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*(4), 667–692. https://doi.org/10.1080/02724980443000142

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341–347.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Breining, B. L., Nozari, N., & Rapp, B. (2019). Learning in complex, multi-component cognitive systems: Different learning challenges within the same system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1093–1106. https://doi.org/10.1037/xlm0000630

Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, *23*(2), 500–506. https://doi.org/10.3758/s13423-015-0900-x

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology*, *97*(3), 210–220. https://doi.org/10.1016/j.ijpsycho.2015.05.004

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, *106*(50), 21442–21446. https://doi.org/10.1073/pnas.0908921106

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 416 – 431. https://doi.org/10.1037/0278-7393.29.3.416

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., & Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Biomedical Engineering*, *53*(12), 2583–2587. https://doi.org/10.1109/TBME.2006.879459

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word production: Behavioral and computational considerations. *The Oxford Handbook of Language Production*, 88–104. https://doi.org/10.1093/oxfordhb/9780199735471.001.0001

De. Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., Huffel, V. S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics*, *8*(2), 135–150. https://doi.org/10.1007/s12021-010-9071-0

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*(4), 725–738. https://doi.org/10.1162/jocn_a_01675

Freund, M. C., Bugg, J. M., & Braver, T. S. (2021). A representational similarity analysis of cognitive control during color-word Stroop. *Journal of Neuroscience*, *41*(35), 7388-7402. https://doi.org/10.1523/JNEUROSCI.2956-20.2021

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (2018). The error-related negativity. *Perspectives on Psychological Science*, *13*(2), 200–204. https://doi.org/10.1177/1745691617715310

Grützmann, R., Endrass, T., Klawohn, J., & Kathmann, N. (2014). Response accuracy rating modulates ERN and Pe amplitudes. *Biological Psychology*, *96*, 1–7. https://doi.org/10.1016/j.biopsycho.2013.10.007

Higby, E., Cahana-Amitay, D., Vogel-Eyny, A., Spiro III, A., Albert, M. L., & Obler, L. K. (2019). The role of executive functions in object-and action-naming among older adults. *Experimental Aging Research*, *45*(4), 306–330. https://doi.org/10.1080/0361073X.2019.1627492

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00255

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

Janssen, N., Carreiras, M., & Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *NeuroImage*, *57*(3), 1243–1250. https://doi.org/10.1016/j.neuroimage.2011.05.015

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*(9), 2960–2969. https://doi.org/10.1093/cercor/bhu092

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General*, *129*(3), 340-360. https://doi.org/10.1037/0096-3445.129.3.340

Korko, M., Coulson, M., Jones, A., & de Mornay Davies, P. (2021). Types of interference and their resolution in monolingual word production. *Acta Psychologica*, *214*, 103251. https://doi.org/10.1016/j.actpsy.2021.103251

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

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*(4), e13990. https://doi.org/10.1111/psyp.13990

Llorens, A., Trébuchon, A., Riès, S., Liégeois-Chauvel, C., & Alario, F.-X. (2014). How familiarization and repetition modulate the picture naming network. *Brain and Language*, *133*, 47–58. https://doi.org/10.1016/j.bandl.2014.03.010

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, *8*.

McDonagh, D., Fisher, A. & Nozari, N. (2020). Do Taxonomic and Associative Relations Affect Word Production in the Same Way? In S. Denison, M. Mack,

Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.* (pp. 544-550). Cognitive Science Society.

Meckler, C., Allain, S., Carbonnell, L., Hasbroucq, T., Burle, B., & Vidal, F. (2011). Executive control and response expectancy: A Laplacian ERP study. *Psychophysiology*, *48*(3), 303–311. https://doi.org/10.1111/j.1469-8986.2010.01077.x

Meckler, C., Carbonnell, L., Ramdani, C., Hasbroucq, T., & Vidal, F. (2017). On-line action monitoring of response execution: An electrophysiological study. *Biological Psychology*, *129*, 178–185. https://doi.org/10.1016/j.biopsycho.2017.09.002

Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early Parallel Activation of Semantics and Phonology in Picture Naming: Evidence from a Multiple Linear Regression MEG Study. *Cerebral Cortex*, *25*(10), 3343–3355. https://doi.org/10.1093/cercor/bhu137

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 65. https://doi.org/10.1037/0278-7393.34.1.65

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. Trends in Cognitive Sciences, 15(10), 453–459. https://doi.org/10.1016/j.tics.2011.07.011

Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a "lexical editor" be? *Journal of Memory and Language*, *60*(2), 291–307. https://doi.org/10.1016/j.jml.2008.09.006

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*(7), 886–903. https://doi.org/10.1080/23273798.2016.1157194

Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics*, *53,* 100875. https://doi.org/10.1016/j.jneuroling.2019.100875

Nuñez, P. L., and Srinivasan, R. (2006). Electric Fields of the Brain: the Neurophysics of EEG. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195050387.001.0001

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*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Oppenheim, G., & Nozari, N. (2021). *Behavioral interference or facilitation does not distinguish between competitive and noncompetitive accounts of lexical selection in word production.* PsyArXiv. https://doi.org/10.31234/osf.io/rjezp

O'Séaghdha, P. G., & Frazer, A. K. (2014). The exception does not rule: Attention constrains form preparation in word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 797–810. https://doi.org/10.1037/a0035576

Python, G., Fargier, R., & Laganaro, M. (2018a). When Wine and Apple Both Help the Production of Grapes: ERP Evidence for Post-lexical Semantic Facilitation in Picture Naming. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00136

Python, G., Fargier, R., & Laganaro, M. (2018b). 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

Qu, Q., Feng, C., & Damian, M. F. (2021). Interference effects of phonological similarity in word production arise from competitive incremental learning. *Cognition*, *212*, 104738. https://doi.org/10.1016/j.cognition.2021.104738

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499. https://doi.org/10.1037/0033-295X.107.3.460

Riès, S., Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., Kuperman, R. A., Auguste, K. I., Brunner, P., Schalk, G., Lin, J. J., Parvizi, J., Crone, N. E., Dronkers, N. F., & Knight, R. T. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences*, *114*(23), E4530–E4538. https://doi.org/10.1073/pnas.1620669114

Riès, S., Janssen, N., Burle, B., & Alario, F.-X. (2013). Response-Locked Brain Dynamics of Word Production. *PLOS ONE*, *8*(3), e58197. https://doi.org/10.1371/journal.pone.0058197

Riès, S., Janssen, N., Dufau, S., Alario, F.-X., & Burle, B. (2011). General-Purpose Monitoring during Speech Production. *Journal of Cognitive Neuroscience*, *23*(6), 1419–1436. https://doi.org/10.1162/jocn.2010.21467

Riès, S. K., Pinet, S., Nozari, N. B., & Knight, R. T. (2021). Characterizing multi-word speech production using event-related potentials. *sychophysiology, 58*(5), e13788. https://doi.org/10.1111/psyp.13788

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (1999). Phonological Segments and Features as Planning Units in Speech Production. *Language and Cognitive Processes*, *14*(2), 173–200. https://doi.org/10.1080/016909699386338

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

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58. https://doi.org/10.1016/j.cogpsych.2013.10.001

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227. https://doi.org/10.1016/j.jml.2005.10.002

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, *106*(1), 322–327. https://doi.org/10.1073/pnas.0805874106

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & Cognition*, *41*(8), 1200–1211. https://doi.org/10.3758/s13421-013-0332-7

Shao, Z., Roelofs, A., Martin, R. C., & Meyer, A. S. (2015). Selective inhibition and naming performance in semantic blocking, picture-word interference, and color–word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1806–1820. https://doi.org/10.1037/a0039363

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *The*

Quarterly Journal of Experimental Psychology, *65*(10), 1927–1944. https://doi.org/10.1080/17470218.2012.670252

Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and prediction error in medial frontal cortex: Integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience*, *5*, 75. https://doi.org/10.3389/fnhum.2011.00075

Strauss, G. P., Allen, D. N., Jorgensen, M. L., & Cramer, S. L. (2005). Test-retest reliability of standard and emotional stroop tasks: An investigation of color-word and picture-word versions. *Assessment*, *12*(3), 330–337. https://doi.org/10.1177/1073191105276375

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*(4), 484–503. https://doi.org/10.1080/23273798.2015.1120878

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011, 879716, *1-13*. https://doi.org/10.1155/2011/879716

Vidal, F., Burle, B., & Hasbroucq, T. (2022). On the Comparison Between the Nc/CRN and the Ne/ERN. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnhum.2021.788167

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the 'error negativity'specific to errors? *Biological Psychology*, *51*(2–3), 109–128. https://doi.org/10.1016/S0301-0511(99)00032-0

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*(5), 575–586. https://doi.org/10.1080/23273798.2017.1395467

Wong, A. W.-K., Chiu, H.-C., Wang, J., Cao, J., Wong, S.-S., & Chen, H.-C. (2017). An early locus of associative and categorical context effects in speech production: Evidence from an ERP study using the picture–word interference paradigm. *Language, Cognition and Neuroscience*, *32*(10), 1305–1319. https://doi.org/10.1080/23273798.2017.1355060

Xue, G., Aron, A. R., & Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, *18*(8), 1923–1932. https://doi.org/10.1093/cercor/bhm220

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological*

Appendix A

Experimental materials.

word	semantic	rhyme	unrelated
cake	pie	rake	bear
pie	cake	tie	hair
rake	hose	cake	tie
nose	hair	hose	pig
hose	rake	nose	wig
hair	nose	bear	pie
pig	bear	wig	nose
bear	pig	hair	cake
tie	wig	pie	rake
wig	tie	pig	hose

Appendix B

Results of Wilcoxon tests and full tables of the MLM analyses for the behavioral data.

Accuracy

Table B1. Wilcoxon Test

	z	p-value
Phase	3.88	< .001
Semantic	-0.12	0.911
Rhyme	0.37	0.715

Table B2. MLM: accuracy ~ phase * condition + (1 | picture) + (1 | subject)

	Coefficient	SE	z	p-value
Intercept	5.48	0.34	16.34	< .001
Phase	-1.53	0.28	-5.44	< .001
Condition (Semantic)	-0.56	0.32	-1.75	0.081
Condition (Phonological)	-0.13	0.35	-0.36	0.716
Phase x Condition (Semantic)	0.66	0.37	1.81	0.071
Phase x Condition (Phonological)	0.40	0.39	1.03	0.306

RT

Table B3. Wilcoxon Test

	z	p-value
Phase	4.70	< .001
Semantic	-0.96	0.347
Rhyme	0.29	0.782

Table B4. MLM: RTlog ~ phase * condition + (1 | picture) + (1 | subject)

	Coefficient	SE	t	p-value
Intercept	-0.53	0.03	-20.33	< .001
Phase	0.14	0.01	25.13	< .001
Condition (Semantic)	0.00	0.01	-0.56	0.577
Condition (Phonological)	0.00	0.01	-0.57	0.566
Phase x Condition (Semantic)	0.00	0.01	-0.39	0.697
Phase x Condition (Phonological)	0.01	0.01	0.89	0.374

Duration

Table B5. Wilcoxon Test

	z	p-value
Phase	-0.77	0.455

Semantic		2.67	0.006
Rhyme		3.49	<.001

Table B6. MLM: durlog ~ phase * condition + (1 | picture) + (1 | subject)

	Coefficient	SE	t	p-value
Intercept	-0.85	0.03	-24.31	< .001
Phase	-0.01	0.00	-1.96	0.050
Condition (Semantic)	0.02	0.00	3.08	0.002
Condition (Phonological)	0.01	0.00	2.94	0.003
Phase x Condition (Semantic)	-0.01	0.01	-0.75	0.452
Phase x Condition (Phonological)	0.02	0.01	2.82	0.005

Appendix C

Figure C1. ERPs for straight (green) and reverse (red) phases for unrelated, semantic, and phonological conditions shown for three electrodes FCz, Cz, and Pz.