## Research report

# Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming

Natalia Shitova [a,b,*], Ardi Roelofs [a], Herbert Schriefers [a],
Marcel Bastiaansen [c,d] and Jan-Mathijs Schoffelen [e]

[a] Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
[b] International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[c] NHTV Breda University of Applied Science, Academy for Leisure, Breda, The Netherlands
[d] Department of Cognitive Neuropsychology, Tilburg School of Social and Behavioural Sciences, Tilburg University, Tilburg, The Netherlands
[e] Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

ABSTRACT

Accumulating evidence suggests that spoken word production requires different amounts of top-down control depending on the prevailing circumstances. For example, during Stroop-like tasks, the interference in response time (RT) is typically larger following congruent trials than following incongruent trials. This effect is called the Gratton effect, and has been taken to reflect top-down control adjustments based on the previous trial type. Such control adjustments have been studied extensively in Stroop and Eriksen flanker tasks (mostly using manual responses), but not in the picture—word interference (PWI) task, which is a workhorse of language production research. In one of the few studies of the Gratton effect in PWI, Van Maanen and Van Rijn (2010) examined the effect in picture naming RTs during dual-task performance. Based on PWI effect differences between dual-task conditions, they argued that the functional locus of the PWI effect differs between post-congruent trials (i.e., locus in perceptual and conceptual encoding) and post-incongruent trials (i.e., locus in word planning). However, the dual-task procedure may have contaminated the results. We therefore performed an electroencephalography (EEG) study on the Gratton effect in a regular PWI task. We observed a PWI effect in the RTs, in

---

* Corresponding author. Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands.
E-mail addresses: n.shitova@donders.ru.nl (N. Shitova), a.roelofs@donders.ru.nl (A. Roelofs), h.schriefers@donders.ru.nl (H. Schriefers), Bastiaansen4.M@nhtv.nl (M. Bastiaansen), jan.schoffelen@donders.ru.nl (J.-M. Schoffelen).

the N400 component of the event-related brain potentials, and in the midfrontal theta power, regardless of the previous trial type. Moreover, the RTs, N400, and theta power reflected the Gratton effect. These results provide evidence that the PWI effect arises at the word planning stage following both congruent and incongruent trials, while the amount of top-down control changes depending on the previous trial type.

## 1.    Introduction

Goal-oriented behavior relies on a flexible system of top-down control that allows for modification of processing strategies based on an analysis of costs and benefits of different processing types in order to improve performance on a task (e.g., Gratton, Coles, & Donchin, 1992). For example, in an experiment with congruent and incongruent stimuli (like in the color-word Stroop task with congruent and incongruent color-word combinations, such as the word *red* in green ink), a participant can adjust the extent to which distractor information is processed depending on the expected trial type (e.g., Lamers & Roelofs, 2011). In particular, participants may choose between a narrow or a wide strategy in stimulus processing depending on whether a more or less thorough analysis is expected to optimize performance. In an experiment, the participant may rely on the previous trial type to form an expectation about the next trial type (Gratton et al., 1992), expecting trial-type repetition (e.g., Egner, 2007). Alternatively, a participant's expectations can be determined by cues that designate the probability of the next trial type (e.g., Aarts & Roelofs, 2011; Aarts, Roelofs, & Van Turennout, 2008; Gratton et al., 1992) or by means of global probability of stimuli of different types (e.g., Carter et al., 2000).

Control adjustments have been intensively studied using Stroop-like tasks. In a seminal study, Gratton et al. (1992) observed that in such tasks, the difference in response time (RT) between incongruent and congruent trials is typically larger following congruent trials than following incongruent trials, referred to as the *Gratton effect* in later research. The Gratton effect has been examined using Stroop and Eriksen flanker tasks (mostly employing manual responses), but not in the picture–word interference (PWI) task (e.g., Glaser & Düngelhoff, 1984), which has been much used in language production research. Recently, Van Maanen and Van Rijn (2010) examined the Gratton effect in PWI embedded in a dual-task paradigm (see below). Aside from reporting (expected) quantitative differences in RTs, they also argued that post-congruent and post-incongruent trials differ with respect to the functional locus of the PWI effect (i.e., the stage at which the incongruence of the picture and the distractor influences processing of the stimulus). We argue that the dual-task procedure might have contaminated the results. The aim of the research reported in the present article was to examine the Gratton effect in regular PWI and to use electroencephalography (EEG) to determine the functional locus of the PWI effect and its modulation by previous trial type.

In the following, we first describe the trial-to-trial manifestation of the Gratton effect in Stroop-like tasks in more detail. Next, we discuss the dual-task findings and locus-shift account of Van Maanen and Van Rijn (2010), and we present an alternative interpretation of their findings in terms of task scheduling rather than a shift in locus. Then, we discuss the results of previous EEG studies on the Gratton effect, which motivate an examination of the Gratton effect in the N400 component of the event-related brain potentials and in frontal theta power. Next, the results of our EEG study are reported. Finally, we discuss the consequences of our findings for the debate about the functional locus of the PWI effect.

### 1.1.    The Gratton effect in RTs

Trial-to-trial sequential effects have been extensively studied using Stroop-like tasks, in which the participant is presented with stimuli that are combinations of a target dimension and a distractor dimension. The participant is instructed to respond to the target dimension while ignoring the distractor dimension. In a proportion of stimuli the target and the distractor dimensions activate the same response (i.e., congruent trials), while in the rest of stimuli the target and the distractor dimensions are associated with different responses (i.e., incongruent trials). For example, in the color-word Stroop task, the stimuli are color words that are printed in a certain color ink (e.g., the word *red* in red or green ink), and the participant is instructed to respond to the color while trying to ignore the word (MacLeod, 1991). In the Eriksen flanker task, the stimuli are strings of letters (e.g., HHSHH or SSSSS) and the participant is instructed to respond to the central (i.e., target) one, while ignoring the distracting flankers (i.e., Eriksen & Eriksen, 1974). The PWI task employs drawings of objects with superimposed object names, and the participant is instructed to name the picture while ignoring the word (Glaser & Düngelhoff, 1984). A common finding of Stroop-like tasks is that participants give slower and less accurate responses on incongruent trials as compared to congruent trials (Eriksen & Eriksen, 1974; Glaser & Düngelhoff, 1984; MacLeod, 1991). Critically, this Stroop-like effect is larger on trials following congruent trials than on trials following incongruent trials (e.g., Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Gratton et al., 1992; Lamers & Roelofs, 2011; Ullsperger, Bylsma, & Botvinick, 2005, for the Eriksen flanker task; Egner & Hirsch, 2005; Kerns et al., 2004; Lamers & Roelofs, 2011; Notebaert, Gevers, Verbruggen, & Liefooghe, 2006, for the Stroop task). The Gratton effect has been taken to reflect adjustments in top-down control that a participant exerts in

certain contexts (but see Hommel, Proctor, & Vu, 2004; Mayr, Awh, & Laurey, 2003).

The Gratton effect has been studied extensively in manual Stroop-like tasks (e.g., Botvinick et al., 1999; Egner & Hirsch, 2005; Kerns et al., 2004) and has been mostly interpreted using the conflict-monitoring account (Botvinick, Braver, Barch, Carter, & Cohen, 2001, but see Aarts et al., 2008; Lamers & Roelofs, 2011). This account suggests that the response conflict that occurs in incongruent trials leads to an increased top-down control during the following trial. The top-down control biases the processing of the stimulus by enhancing the target stimulus dimension or suppressing the irrelevant distractor dimension, thereby diminishing the magnitude of the Stroop-like effect.

Research on the Gratton effect in speech production paradigms has remained scarce (Lamers & Roelofs, 2011, for the Eriksen flanker task with vocal responding and the Stroop task; Freund, Gordon, & Nozari, 2016, and Van Maanen & Van Rijn, 2010, for the PWI task). Van Maanen and Van Rijn investigated the control adjustments using a dual-task procedure (Pashler, 1984). The participants had to perform two tasks at the same time: a tone classification task and a PWI task. In the PWI task, picture and distractor word were semantically related (e.g., a picture of a cat combined with the word *dog*, the incongruent condition), semantically unrelated (e.g., cat − *house*, the unrelated condition), or the distractor word was the name of the picture (cat − *cat*, the congruent condition). Critically, the interval between the tone and the PWI stimulus (stimulus-onset asynchrony − SOA) was manipulated. The SOAs were 100, 350, and 800 msec. The overall naming RTs decreased with increasing SOA. Moreover, Van Maanen and Van Rijn suggested that at the longest SOA (i.e., 800 msec) the semantic interference effect (i.e., the longer RTs observed for semantically related compared with unrelated trials) was smaller on trials following incongruent trials than following congruent trials, reflecting the Gratton effect. At the shorter SOAs, the semantic interference effect was present on trials following incongruent trials, but the semantic effect was absent on trials following congruent trials. Under the assumption of a response-selection bottleneck (Pashler, 1984), the absence of the semantic effect at short SOAs suggests that the effect arises before response selection in perceptual and conceptual encoding, whereas full-blown semantic interference suggests a locus in response selection or later (for an extensive discussion, see Piai, Roelofs, & Schriefers, 2014). Following congruent trials, the semantic effect was absent at short SOAs, suggesting a locus in perceptual and conceptual encoding, whereas following incongruent trials, the semantic effect was present, suggesting a locus in word planning or later. According to Van Maanen and Van Rijn, these findings indicate a shift in the functional locus of PWI from perceptual and conceptual encoding on post-congruent trials to word planning on post-incongruent trials.

It is important to note that researchers have found no agreement on the functional locus of interference in the PWI task. Picture naming is generally assumed to consist of three main stages (e.g., Levelt, Roelofs, & Meyer, 1999; Roelofs, 2003, 2014). First, the picture is perceived and conceptually identified based on the stimulus features, hereafter the *perceptual and conceptual encoding* stage. According to a meta-analysis of Indefrey and Levelt (2004; Indefrey, 2011), this stage is completed around 200 msec post stimulus-onset. Then, the corresponding spoken word is planned based on information retrieved from long-term memory, hereafter the *word planning* stage. This stage starts about 200 msec post stimulus-onset and ends around 150 msec before articulation onset. Finally, the planned picture name is articulated. As we discussed in detail elsewhere (Shitova, Roelofs, Schriefers, Bastiaansen, & Schoffelen, 2016), the locus of the interference effect has been linked either to perceptual and conceptual encoding (Dell'Acqua, Job, Peressotti, & Pascali, 2007; Dell'Acqua et al., 2010; Van Maanen, Van Rijn, & Borst, 2009) or to word planning (Piai, Roelofs, & Schriefers, 2014; Schnur & Martin, 2012). The dual-task findings by Van Maanen and Van Rijn (2010) would suggest that the locus is in perceptual and conceptual encoding or word planning depending on the previous trial type.

Van Maanen and Van Rijn (2010) demonstrated the utility of their locus-shift account in computer simulations using the RACE/A model. The Gratton effect was instantiated in the model as a modulation of the speed of processing of words relative to pictures. Assuming that the participants exert more top-down control (i.e., "more suppression of the reading response", p. 175), on post-incongruent trials than on post-congruent trials, the speed of word processing was set lower for post-incongruent than for post-congruent trials in the computer simulations. As a consequence, the locus of the semantic effect was in perceptual and conceptual encoding for post-congruent trials but in word planning for post-incongruent trials in the model.

However, there is an alternative interpretation of the empirical results reported in the Van Maanen and Van Rijn (2010) study, which does not assume that the locus of the interference effect shifts depending on the previous trial type. Rather, the alternative interpretation holds that there is a single fixed locus of semantic and Stroop-like effects (i.e., in word planning, see Piai, Roelofs, & Schriefers, 2014; Shitova et al., 2016). The behavioral effects reported by Van Maanen and Van Rijn may also have resulted from a change in strategic task scheduling depending on the previous trial type, in order to optimally perform in the dual-task setting. Piai, Roelofs, and Schriefers (2014) argued that participants may strategically control the amount of overlap between response-selection processes in the two tasks during dual-task performance. On post-congruent trials, the participants of Van Maanen and Van Rijn might have adopted a more daring strategy, allowing overlap in response-selection processes between the tone and PWI tasks. In this way, word planning for the PWI task could have started before the response for the tone classification task was selected, such that the semantic PWI effect overlaid the SOA effect and these effects were underadditive on post-congruent trials. However, on post-incongruent trials, the participants might have adopted a more conservative strategy, not allowing such overlap, i.e., word planning for the PWI task only started after the response for the tone classification task was selected, which made the SOA effect and the semantic effect additive on post-incongruent trials. Thus, this difference in task scheduling can explain the additivity of the SOA and the semantic effects

on post-incongruent trials and the underadditivity on post-congruent trials, which was observed by Van Maanen and Van Rijn. Thus, the difference in semantic effect at short SOAs between post-congruent and post-incongruent trials does not need to reflect a shift in locus of the PWI effect, as Van Maanen and Van Rijn maintain, but may instead reflect a difference in strategic task scheduling.

If instead the previous trial type leads to a shift of the functional locus, as suggested by Van Maanen and Van Rijn (2010), this may have important implications for the interpretation of previous studies examining the functional locus of PWI effects. In such studies, congruent and incongruent trials are presented to participants in pseudo-randomized order and the critical measures are usually obtained through averaging trials within conditions irrespective of the previous trial type. However, if the locus of interference systematically shifts following congruent and incongruent trials the results of such averaging depend on the proportions of the transitions of each type, which are not routinely controlled. Thus, the question of whether the locus of interference in the PWI task indeed depends on the previous trial type is important, yet it needs to be investigated without a potentially confounding dual-task paradigm, which may contaminate the results as we indicated. Therefore, in the present study we investigated the behavioral and electrophysiological correlates of the Gratton effect in a regular PWI task.

## 1.2. The Gratton effect in electrophysiological measures

Electrophysiological correlates of control adjustment effects have been observed in a number of Stroop-like tasks. A typical finding for the Eriksen flanker task is a larger N2 component (i.e., "second negative peak in the averaged ERP waveform", Folstein & Van Petten, 2008) over fronto-central electrode sites on incongruent compared to congruent trials (Kopp, Rist, & Mattler, 1996). This N2 effect was consistently smaller following incongruent than following congruent trials (Clayson & Larson, 2011; Larson, Clayson, & Clawson, 2014, for a review). For the Stroop task (in various adaptations), a prominent negative-going deflection at approximately 400—450 msec post-stimulus onset, an N400 component, is typically larger on incongruent than on congruent trials (Hanslmayr et al., 2008; Liotti, Woldorff, Perez, & Mayberg, 2000). Some of these previous studies referred to this component as an "N450", but we take it to be an N400 (see Piai, Roelofs, Jensen, Schoffelen, & Bonnefond, 2014; Piai, Roelofs, & Van der Meij, 2012; Roelofs, Piai, Garrido Rodriguez, & Chwilla, 2016; Shitova et al., 2016). Contrary to the N2 effect in the Eriksen flanker task, the N400 effect has not been consistently shown to follow the Gratton pattern of control adjustments in the Stroop task. Null results were reported for a manual color-word Stroop task (Larson, Clawson, Clayson, & South, 2012; Larson, Kaufman, & Perlstein, 2009) and an auditory Stroop task with manual, overt vocal, and covert vocal responses (Donohue, Liotti, Perez, & Woldorff, 2012). West, Bailey, Tiernan, Boonsuk, and Gilbert (2012), however, reported that the N400 effect was larger following congruent trials than following incongruent trials in a counting Stroop task (for similar effects of global probability on control adjustments in the counting

Stroop task, see West & Bailey, 2012). Moreover, Larson, Clayson, Kirwan, and Weissman (2016) reported control adjustment effects on the N400 component in a novel prime-probe Stroop-like task. The N400 effect has also been reported to reflect the semantic and Stroop-like effects in the PWI task (e.g., Piai, Roelofs, Jensen, et al., 2014; Piai et al., 2012; Shitova et al., 2016). Based on the N400 findings, it has been argued that the semantic and Stroop-like effects in the PWI task arise during word planning rather than during perceptual and conceptual encoding. As indicated before, the perceptual and conceptual encoding stage is completed around 200 msec post stimulus-onset, which is well before the moment that the Stroop-like effects occurs in the event-related potentials (ERPs). For example, in the study of Shitova et al. (2016), the onset of the Stroop-like effect was around 350 msec after stimulus onset. This indicates that the effect occurred during word planning, which starts 200 msec post stimulus-onset. However, the N400 effect has never been used for studying control adjustments in speaking.

As concerns induced patterns of electrophysiological activity, power changes in the theta-band (4—8 Hz) from around 400 msec post-stimulus until response onset over medial frontal cortex have consistently been reported for Stroop-like tasks (e.g., Hanslmayr et al., 2008; Nigbur, Ivanova, & Stürmer, 2011; Pastötter, Dreisbach, & Bäuml, 2013) as well as for feedback and error processing (Cavanagh, Cohen, & Allen, 2009; Cavanagh, Zambrano-Vazquez, & Allen, 2012). Interestingly, Hanslmayr et al. showed for the manual Stroop task that theta-band power over the anterior cingulate cortex (ACC) increased monotonically from trials that yielded the least conflict (i.e., congruent) to trials that yielded the most conflict (i.e., incongruent trials for which the target dimension matched the distractor dimension on the previous trial). Thus, mid-frontal theta power reflected not only the presence of conflict, but also its degree, which makes this frequency band interesting for research on control adjustments. Midfrontal theta power effects have only been reported once for PWI: Piai, Roelofs, Jensen, et al. (2014) demonstrated that theta-band power over left superior frontal gyrus was larger on incongruent than congruent trials in the PWI task, reflecting competitive selection of words in picture naming.

To summarize, researchers have found no agreement on the functional locus of the Gratton effect in picture naming: perceptual and conceptual encoding versus word planning. However, only a few studies (Freund et al., 2016; Van Maanen & Van Rijn, 2010) have examined the Gratton effect in PWI. Therefore, we conducted a new study examining the Gratton effect in standard PWI and measured ERPs and oscillatory power to obtain evidence on the functional locus of the effect. In their meta-analysis, Indefrey and Levelt (2004; Indefrey, 2011) estimated that perceptual and conceptual encoding in picture naming is completed at about 200 msec after picture onset, after which word planning starts. If the hypothesis of Van Maanen and Van Rijn of a differential locus is correct, the Stroop-like effect in PWI should appear in different time windows depending on the previous trial type, i.e., during perceptual and conceptual encoding on post-congruent trials (i.e., before 200 msec post-stimulus onset) or during word planning on post-incongruent trials (i.e., after 200 msec post-

stimulus onset). However, if the functional locus of the Stroop-like effect does not depend on the previous trial type, it is expected to appear during word planning on both post-congruent and post-incongruent trials.

## 2. Method

### 2.1. Ethics statement

The line of research within which the study was performed was approved by the local Ethics Committee (ECG-2013-2504-095-102) and followed the World Medical Association Declaration of Helsinki. An informed consent form was obtained from all participants before the experiment.

### 2.2. Participants

Twenty-five participants (19 to 30 years old, mean 22.3 years, ten male) took part in the study for course credits or monetary compensation. All participants were right-handed, native speakers of Dutch, with normal or corrected to normal vision, and reported no history of neurological disorders.

### 2.3. Stimuli and procedure

The participants were seated approximately 70 cm away from a computer screen. The participants were instructed to name the picture as fast and accurately as possible, while ignoring the distractor word. They were also informed that the stimuli would only appear in the center of the screen. They first familiarized themselves with the stimulus set using a printout of all pictures with names used in the experiment. Then the participants performed a practice session of ten trials. Up to four additional practice blocks of 10 trials each were performed by the participants if they requested more practice or if the experimenter was not satisfied with the performance on the previous block(s). Following the training, the participants performed five blocks of the task (156 trials per block). Small breaks (up to 5 min) were offered to the participants between blocks. The PWI stimuli consisted of a line drawing (4 × 4 cm) with a superimposed distractor word printed in lower-case Arial 24pt letters (1.8−3.2 cm wide, .5−.7 cm high).

A trial is described in Fig. 1. It started with the presentation of four asterisks for 1200 msec during which the participants



**Fig. 1 − Trial events.**

were instructed to blink. After that a blank screen appeared for a random time between 700 and 1000 msec. The stimulus appeared immediately after that and stayed on the screen for 700 msec, after which the screen remained blank for 800 msec. Trials with response latencies shorter than 300 msec or longer than 1500 msec were excluded from the analysis. An experimental session including EEG application and instruction took on average a little more than two hours, while the task alone took approximately 50 min. The experiment was programmed using Presentation software (NeuroBehavioral Systems).

The stimulus set was composed of 39 objects (13 groups of three items each). Pictures of objects were paired with distractor words to create congruent and incongruent stimuli: For a congruent stimulus a picture of an object was paired with its name, and for an incongruent stimulus a picture of an object was paired with the name of one of the other two objects from the same semantic domain (see Appendix A for stimulus materials). The pictures were line drawings from a picture database of the Max Planck Institute for Psycholinguistics, from open source internet resources, or drawn from scratch. The pictures were previously used in a few naming experiments at our lab and showed high name agreement. The picture-word pairings were also used in our previous PWI study (Shitova et al., 2016) and showed a highly reliable Stroop-like effect.

A special pseudo-randomization technique was employed in order to create individual stimulus lists for each participant. The stimuli were pseudo-randomized using Mix software (Van Casteren & Davis, 2006) such that two stimuli that belonged to the same semantic group were separated by at least two other stimuli, the condition of the stimuli could not repeat more than three times in a row, and the first letter of two neighboring stimuli could not be the same. The maximal difference between the total number of trials of each of the four types, i.e., congruent-after-congruent, congruent-after-incongruent, incongruent-after-congruent, and incongruent-after-incongruent (for each participant), was less than or equal to 8.

### 2.4. Recording

The participant's responses were recorded for offline semi-automatic RT estimation using PRAAT (Boersma, 2002). Naming errors were marked by the experimenter during the task. After per-participant and per-condition proportions of errors were calculated, error trials and post-error trials were discarded from analysis, both behavioral and EEG.

A standard 10−20 64-channel ActiCAP system (Brain Products) was used for scalp EEG measurements. The data were referenced online to FCz and re-referenced offline to the average of left and right mastoids. Four additional pairs of bipolar electrodes registered eye and lip movements, as well as the electrocardiogram (ECG). For the vertical electrooculogram (EOG) two electrodes were placed above and below the left eye. For the horizontal EOG two electrodes were placed on the left and right temples. For the lip electromyogram (EMG) two electrodes were placed at the left orbicularis oris superior and the right orbicularis oris inferior. For the ECG two electrodes were placed on the left and right collar bones. The ECG data were not used for the analysis reported in this article. The
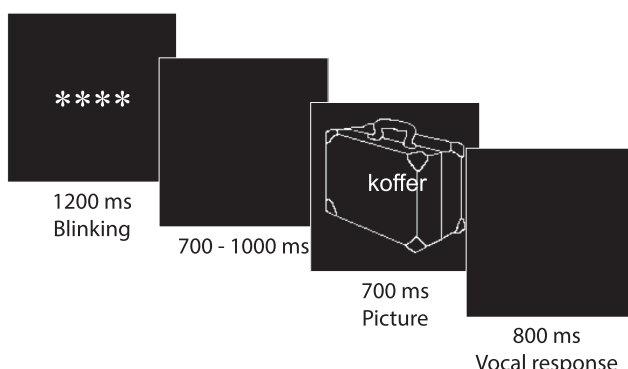
signals were digitized at a sampling rate of 1000 Hz and online-filtered with the low cutoff of .016 Hz and the high cutoff of 125 Hz.

## 2.5. Analysis

The data were analyzed with FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom analysis scripts using Matlab v.8.1.0.604 (R2013a, The MathWorks, Inc.). Epochs started at 500 msec before the onset of the stimulus and ended at the overt articulation onset (so, the length of trials was variable). The data were further inspected for EOG artifacts (blinks, eye movements) and artifactual trials were removed from the analysis. Then we inspected the data for muscular artifacts associated with speech onset in a manual trial-by-trial procedure. Such artifacts typically manifested as high-frequency noise over a spatially widespread set of EEG channels, starting at about 100—200 msec before the speech onset. If such artifact appeared earlier than 300 msec after the stimulus onset the trial was discarded completely, if the artifact occurred later, only the contaminated part of the trial was removed. As a result of this procedure the EEG trials were of different lengths. All other types of artifacts including body movements and facial muscle artifacts not associated with speech were also discarded from the analysis via visual inspection.

Due to blinks, a median of 3.6% of the data were rejected (Q1 = 1.9%, Q3 = 5.7%). Through visual inspection, a median of .38% of the data were rejected (Q1 = 0, Q3 = 1.9%). Additionally, due to errors in naming (both on the current trial and on the previous trial), a median of 5.3% of the data were rejected (Q1 = 3.8%, Q3 = 8.0%). In total, a median of 10.5% of the data were excluded from analysis (Q1 = 6.8%, Q3 = 16.1%).

The data of one participant were discarded from the analysis due to extremely poor performance: The RTs were more than 2.5 standard deviations longer than the grand average RTs. The data of one more participant were discarded due to excessive blinking rate (almost 20% of the data) and residual eye movements that contaminated the rest of the data. This resulted in using the data of 23 participants for all types of analysis. Furthermore, electrode sites T7 and T8 were discarded from analysis due to high levels of noise in almost all participants' data.

For the ERP analysis the data were further band-pass filtered at .5—40 Hz with a onepass-zerophase linear non-causal hamming-windowed finite impulse response (FIR) filter and baseline-corrected using the last 300 msec of the pre-stimulus interval. The ERPs were computed through averaging epochs per condition and per participant. We created 8 condition-specific sets of ERPs per participant: current-congruent, current-incongruent, previous-congruent, previous-incongruent, congruent-after-congruent, congruent-after-incongruent, incongruent-after-congruent, incongruent-after-incongruent. Due to the difference in length between trials, the individual ERPs for the different conditions and participants also differed in length. Moreover, such individual ERPs were composed by diminishing number of trials as a function of time (since the RT differed across trials, see Appendix B).

In order to obtain the time—frequency representation (TFR) of the data, a sliding window Fourier Transform was used. We computed time-resolved power for frequencies within the 1—40 Hz band, using the entire time-window from 700 msec before stimulus onset until the end of the trial in steps of 10 msec. A variable length Hanning-tapered window was applied to estimate the power at each frequency using 3 oscillation cycles. Single trial TFRs were averaged condition-specifically for each participant analogous to the ERPs. These individual TFRs were further baseline-corrected using a pre-stimulus interval from 300 msec before stimulus onset until stimulus onset, averaged over trials of all conditions. For the baseline correction we first calculated baseline activity for every frequency, and then performed a relative change correction, i.e., subtraction of baseline activity values from post-stimulus power estimates with further division by the same baseline values.

The Stroop-like effect in ERPs and TFRs was assessed through contrasting incongruent-after-congruent trials with congruent-after-congruent trials (i.e., a post-congruent Stroop-like effect), and incongruent-after-incongruent trials with congruent-after-incongruent (i.e., a post-incongruent Stroop-like effect). The Gratton effect in ERPs and TFRs was assessed through contrasting difference ERPs and TFRs corresponding to post-congruent and post-incongruent Stroop-like effects.

Statistical significance of the difference between conditions in ERPs and TFRs was assessed using a cluster-based permutation approach (Maris & Oostenveld, 2007). For every contrast, a paired-samples $t$-statistic was calculated per electrode and time-point. Subsequently, these spatio-temporal maps were thresholded at a nominal $p$-value of .05 (two-sided). Next, clusters were created from the thresholded $t$-maps, where adjacent electrode and time-points were clustered, and the cluster statistic was calculated as the sum of the $t$-values within the cluster. The maximum cluster statistic was then compared against a permutation distribution, which was constructed through random re-labeling of the conditions of the original condition-specific ERPs or TFRs for 1000 times and determining spatio-temporal clusters as described earlier. Under the null hypothesis of exchangeability of ERPs or TFRs between conditions, the percentage of permutations that returned a larger cluster statistic than the original cluster statistic served a cluster-based $p$-value. The critical alpha-level for a one-sided and a two-sided test were put at .05 and .025, respectively. In order to assess statistical significance of the Stroop-like effects in ERPs we applied a two-sided test over the entire time-line (i.e., starting 500 msec before the stimulus onset and ending at the last time-point when individual ERPs for all participants and both conditions were available). For the analysis of the Gratton effect within the N400 time-window, the time-line was limited to 450—500 msec post-stimulus onset and a one-sided test was applied, since the difference between incongruent and congruent trials was expected to be larger following congruent rather than incongruent trials. Positive-sided tests were applied to TFRs. For the analysis of the Stroop-like effects the theta power was averaged per participant per condition (4—6 Hz) within

the time-window of 350–650 msec post-stimulus onset, and these estimates entered the cluster-based permutation analysis. The Gratton effect was tested within the time-window of 350–550 msec post-stimulus onset.

## 3.     Results

### 3.1.     Behavioral data

The group average RTs and error rates are presented in Fig. 2. The figure shows that in both RTs and error rates the Stroop-like effect was smaller following incongruent trials than following congruent trials. In the RTs, the interaction was driven by the current-congruent trials (i.e., longer RTs on the congruent-after-incongruent trials than on the congruent-after-congruent trials), whereas it was driven by the current-incongruent trials in the error rates (i.e., lower error rates on the incongruent-after-incongruent trials than on the incongruent-after-congruent trials).

For the RT data, a 2 × 2 repeated measures analysis of variance (ANOVA) with the factors current trial (congruent, incongruent) and previous trial (congruent, incongruent) showed significant main effects of current trial [$F(1,22) = 211.64, p < .001$], previous trial [$F(1,22) = 9.85, p = .005$], and an interaction between current trial and previous trial [$F(1,22) = 23.20, p < .001$]. RTs in current-congruent trials were slowed down if the previous trial was incongruent compared to congruent [$t(22) = -5.18, p < .001$]. However, there was no influence of the previous trial type on RTs in current-incongruent trials [$t(22) = 1.13, p = .27$].

For the accuracy data, a 2 × 2 repeated measures ANOVA with the factors current trial (congruent, incongruent) and previous trial (congruent, incongruent) showed significant main effects of current trial [$F(1,22) = 48.41, p < .001$], previous trial [$F(1,22) = 16.24, p < .001$], and an interaction between current trial and previous trial [$F(1,22) = 10.38, p = .004$]. A further pairwise t-test showed that error rates were reduced in current-incongruent trials if the previous trial was incongruent compared to congruent [$t(22) = 3.65, p = .001$]. This effect was absent in current-congruent trials [$t(22) = .46, p = .64$]. The interaction reflected that the Stroop-like effect (i.e., the difference between incongruent and congruent trials) was larger following congruent than following incongruent trials in both RTs and error rates, hence replicating the Gratton effect.

### 3.2.     ERP data

Group average ERPs on correct trials only for the four current trial by previous trial conditions over FCz are shown in Fig. 3 (A; for group average ERPs over nine standard electrode sites see Appendix C). Inspecting the curves visually, it can be observed that the waveforms deflected more negatively in the incongruent conditions than in the congruent conditions during the N400 time-window (approximately from 350 to 600 msec after the stimulus presentation). Moreover, as depicted in Fig. 3(B), the difference between the incongruent and congruent conditions (i.e., the Stroop-like effect) was larger following congruent than following incongruent trials, which is the Gratton effect.

A two-sided cluster-based permutation test performed on all time-points and all electrode sites yielded significant differences between congruent and incongruent stimuli (a Stroop-like effect) on both post-congruent and post-incongruent trials. Following congruent trials, the Stroop-like effect was present over a fronto-central cluster that was observed from approximately 360 until 760 msec post-
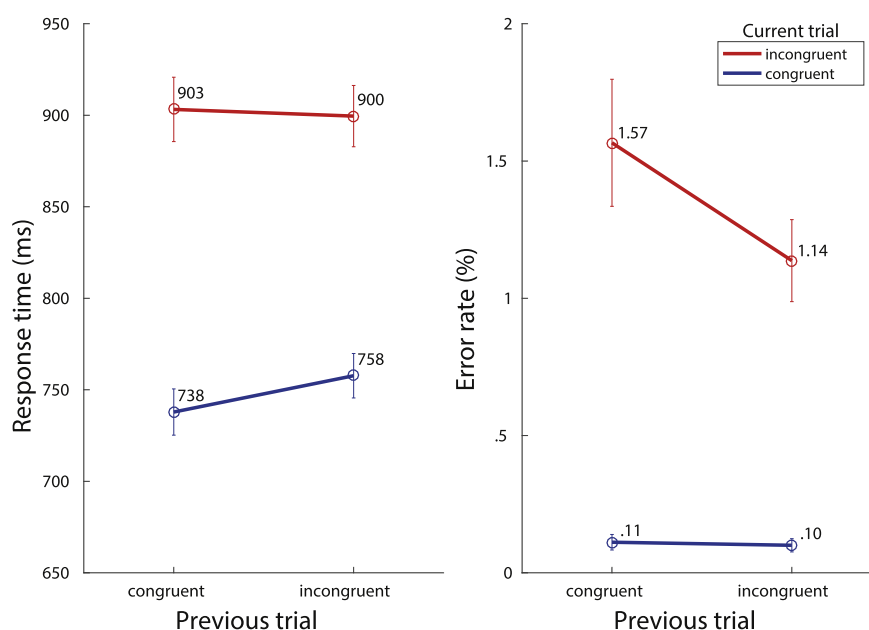


Fig. 2 – Mean response time and error rate as a function of previous trial (congruent, incongruent) and current trial (congruent, incongruent). The error bars indicate the standard error of the mean per condition.
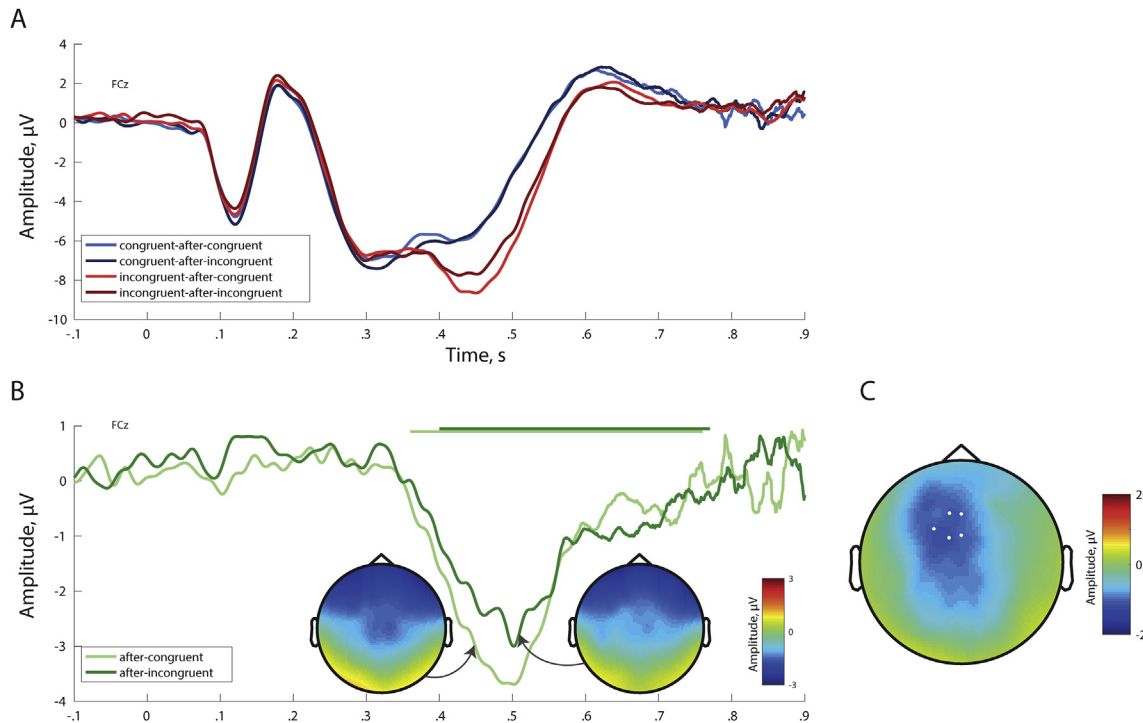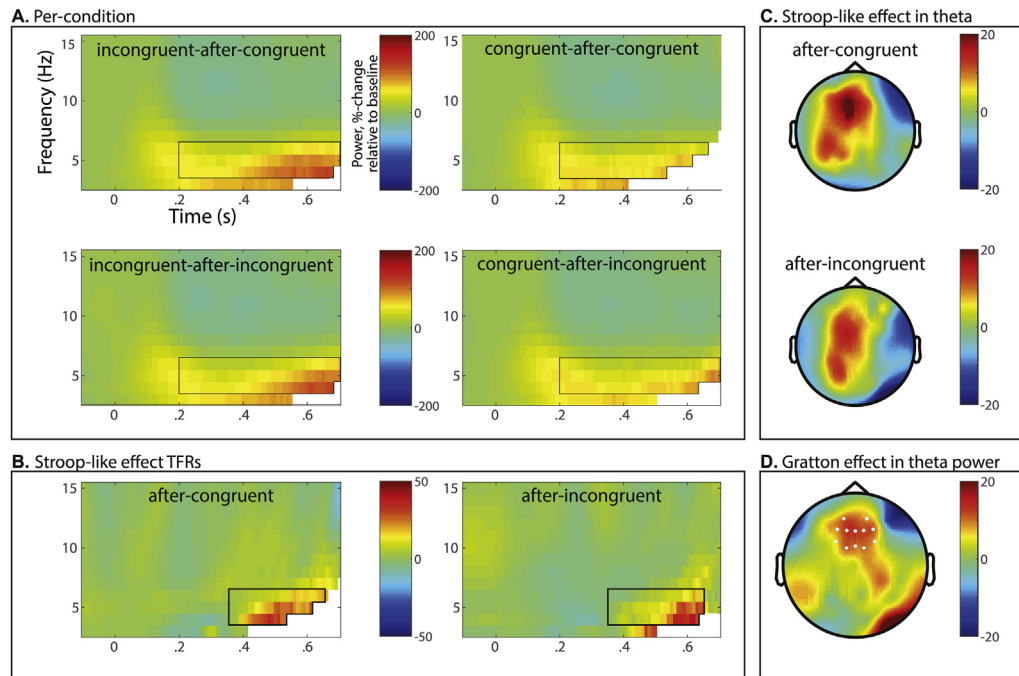
Fig. 3 — The N400 effects. A. Group-average sensor-level ERPs corresponding to the congruent-after-congruent, congruent-after-incongruent, incongruent-after-congruent, and incongruent-after-incongruent conditions, shown for FCz. B. Difference waveforms for after-congruent and after-incongruent trials, calculated as current-incongruent ERP minus current-congruent ERP. Horizontal lines mark the time-windows during which the Stroop-like N400 effects were significant in after-congruent and after-incongruent trials. Topographies of the Stroop-like effect in the N400 on after-congruent (left) and after-incongruent (right) trials, calculated through averaging amplitudes of the difference group ERPs within the time-windows in which the difference between current-congruent and current-incongruent conditions was significant. C. Topography of the Gratton effect, calculated through averaging amplitudes of the difference of the difference group ERPs within 450–500 msec post-stimulus onset. The highlighted electrode sites (F1, Fz, FC3, FC1, FCz) are those that entered the spatio-temporal cluster based on which we rejected the null hypothesis of equal magnitudes of the Stroop-like N400 effect in post-congruent and post-incongruent trials.

stimulus onset ($p < .001$). Following incongruent trials, the Stroop-like effect emerged in a similarly distributed spatio-temporal cluster from approximately 400 until 770 msec post-stimulus onset ($p < .001$, see topographies of the Stroop-like N400 effects in Fig. 3B).

Visual inspection of difference waveforms suggested that the post-congruent and post-incongruent difference wave-forms deflected differentially within the time interval of the N400 component (i.e., from approximately 350 to 600 msec post-stimulus onset) with the maximum difference reached within 450–500 msec post-stimulus onset over a range of midcentral electrode sites. Comparison of topographies of the Stroop-like effect after congruent and after incongruent trials revealed a difference across fronto-central electrode sites, slightly left-lateralized, centered around FC1 (see Fig. 3C). A cluster-based permutation test showed a signifi-cant interaction between current-trial and previous-trial factors over the cluster of five fronto-central electrode sites (F1, Fz, FC3, FC1, FCz; $p = .02$): The negative-going difference between the current-incongruent and current-congruent

trials within the 450–500 msec post-stimulus onset was larger following congruent trials than following incongruent trials, reflecting the Gratton effect in the N400. A post-hoc analysis of simple effects showed that within the time-window 450–500 msec post-stimulus onset the incongruent-after-congruent trials were associated with a larger negative deflection (i.e., the N400 component) than the incongruent-after-incongruent trials ($p = .015$), while no difference was observed between congruent-after-congruent and congruent-after-incongruent trials. Visual inspection of the distribu-tions of peak latencies showed no systematic shift, and a two-sided $t$-test did not reveal difference between N400 peak latencies of the post-congruent and post-incongruent trials [$t(22) = -.52$, $p = .61$, CI $= (-23.9, 14.4)$]. A two-sided Bayesian $t$-test confirmed absence of systematic difference in timing of the N400 ($BF_{01} = 4.05$, suggesting that the data were ~4 times more likely to be observed under the null hypothesis of no difference between conditions than under the alternative hypothesis of systematic shift in timing of the N400).

**Fig. 4 — The theta-power effects. A. Group-average sensor-level TFRs corresponding to the incongruent-after-congruent, incongruent-after-incongruent, congruent-after-congruent, and congruent-after-incongruent conditions, averaged over 11 midcentral sites: AF3, F5, F3, F1, Fz, FC3, FC1, FCz, C3, C1, Cz. B. The Stroop-like theta power effect (i.e., the difference between the current-incongruent and current-congruent TFRs) for after-congruent (left) and after-incongruent (right) conditions. C. The Stroop-like theta power effect topographies for after-congruent (top) and after-incongruent (bottom) conditions. The topographical distributions were calculated through averaging power of the difference group TFRs between 350 and 650 msec post-stimulus onset. D. The Gratton theta-power effect topography. The topographical distribution of the Gratton effect was calculated through averaging power of the difference of the difference group TFRs between 350 and 550 msec post-stimulus onset. The highlighted electrode sites are those that entered the spatio-temporal cluster based on which we rejected the null hypothesis of equal magnitudes of the Stroop-like theta power effect in post-congruent and post-incongruent trials.**

### 3.3.    Time–frequency data

Fig. 4 shows the time–frequency data (TFR). Visual inspection of group-averaged TFRs revealed theta-power (4–6 Hz) increase relative to the pre-stimulus baseline in all conditions, centered around midfrontal electrode sites (F1, Fz) and persisting from approximately 200 msec post-stimulus onset until the end of the trial.

Importantly, this event-related synchronization was larger on incongruent trials compared to congruent trials from approximately 350 to 650 msec post-stimulus onset (see Fig. 4 for TFRs and topographies). This Stroop-like effect was significant in the cluster-based permutation test that included all electrode sites and all time-points between 350 and 650 msec post-stimulus onset ($p = .03$ for the post-congruent comparison; $p = .04$ for the post-incongruent comparison). Moreover, the Stroop-like effect in the midfrontal theta power was larger following congruent than incongruent trials between 350 and 550 msec post-stimulus onset, reflecting the Gratton effect in the theta power ($p = .046$).

### 4.    Discussion

Researchers have found no agreement on the functional locus of the Gratton effect in picture naming: perceptual and conceptual encoding versus word planning. Above, we reported the results of an EEG study on the Gratton effect in regular PWI. Previous work indicates that the stage of perceptual and conceptual encoding is completed around 200 msec post stimulus-onset, followed by the stage of word planning that lasts until around 150 msec before articulation onset. We observed that the PWI effect was present in the RTs, N400, and frontal theta power, regardless of previous trial type. Moreover, the RTs, N400, and theta power reflected the Gratton effect. The onset of the effects was about 350–400 msec post stimulus-onset, which is well after the completion of the stage of perceptual and conceptual encoding. Instead, the timing of the effects suggests that they arise during word planning. Below, we argue that these results challenge the locus-shift account of Van Maanen and Van Rijn. Our results indicate

that the PWI effect arises at the word planning stage following both congruent and incongruent trials, while the amount of top-down control changes depending on the previous trial type.

### 4.1. Challenge to the locus-shift account

The only previous study that addressed the question of the locus of the Gratton effect in PWI (Van Maanen & Van Rijn, 2010) used a dual-task procedure. On long-SOA trials, the Gratton effect manifested itself as a numerical (but not statistically significant; see our reanalysis of their data below) reduction of the semantic effect on the trials following incongruent trials as compared to the trials following congruent trials. Furthermore, on short-SOA trials, the semantic effect vanished following congruent trials but remained following incongruent trials. Van Maanen and Van Rijn interpreted this pattern of results as evidence for a shift in the functional locus of the semantic interference effect: Early during processing (i.e., in perceptual and conceptual encoding) following congruent trials and late (i.e., in word planning) following incongruent trials. We argued that the pattern of short-SOA results by Van Maanen and Van Rijn could be explained differently, in terms of a task scheduling difference between post-congruent and post-incongruent trials. Following congruent trials, the participants might have adopted a daring strategy that allowed overlap in response-selection processes related to the two tasks, which led to the absence of semantic interference at short SOAs. However, following incongruent trials, the participants might have used a conservative strategy that allowed little overlap, thus the semantic effect was present at short SOAs. Therefore, the difference in semantic effect at short SOAs between post-congruent and post-incongruent trials does not need to reflect a shift in locus of the PWI effect, as Van Maanen and Van Rijn maintain, but may instead reflect a difference in strategic task scheduling.

We therefore used a standard PWI task and recorded EEG, as well as behavioral data. We replicated the behavioral findings with respect to the Gratton effect (Freund et al., 2016): Both the RTs and the error rates showed that the Stroop-like effect diminished on the trials following incongruent trials as compared to the trials following congruent trials. As extensively discussed by Lamers and Roelofs (2011), studies in the literature show that the Gratton effect may be driven by current incongruent trials (Kerns et al., 2004), current congruent trials (Lamers & Roelofs, 2011), or both (Gratton et al., 1992). To explain this difference among studies in the underlying RT patterns giving rise to the Gratton effect (i.e., a larger Stroop-like effect on post-congruent than on post-incongruent trials), Lamers and Roelofs presented a two-factor account. According to this account, a difference in the extent to which distractor information is processed between post-congruent and post-incongruent trials gives rise to the basic Gratton effect, where a post-incongruent cost for congruent trials and a post-incongruent benefit for incongruent trials are observed.

In addition, response caution is assumed to differ between post-congruent and post-incongruent trials. The caution is higher following post-incongruent trials equally prolonging RTs in the congruent and the incongruent condition on post-incongruent trials relative to post-congruent trials. In this way, the basic RT patterns underlying the Gratton effect are modulated, yielding the RT patterns observed by Kerns et al. (2004), Lamers and Roelofs (2011), and Gratton et al. (1992). In the present experiment, we observed the RT pattern obtained by Lamers and Roelofs. Greater caution on post-incongruent trials is likely to diminish error rates, therefore increasing the post-incongruent benefit for current-incongruent trials, as we saw in the present experiment. Lamers and Roelofs presented the results of computer simulations using the WEAVER++ model to show the utility of their two-factor account.

Moreover, we investigated the electrophysiological correlates of the Stroop-like effect (i.e., the N400 effect and the midfrontal theta power). The results paralleled the behavioral findings: The Stroop-like effect was present in the N400 component and the theta power regardless of the previous trial type, but it diminished after incongruent trials. No earlier effects were observed when comparing current congruent and current incongruent trials following either congruent or incongruent trials. Thus, the Gratton effect manifested itself only quantitatively (in RTs, N400 amplitude, and theta power) and we found no evidence for a shift of the locus of the Stroop-like effect depending on the previous trial type. Therefore, our findings challenge the conclusions of Van Maanen and Van Rijn.

Apart from design differences between the experiment by Van Maanen and Van Rijn (2010) and our experiment (i.e., a PWI task embedded in a dual-task paradigm vs the standard PWI task), the studies employed different experimental contrasts to assess the Gratton effect in picture naming. In our study we focused on the Stroop-like effect (i.e., congruent vs incongruent trials), while Van Maanen and Van Rijn studied the semantic effect (i.e., neutral vs incongruent trials, also called unrelated vs semantically related trials). Note that in the PWI task, the Stroop-like contrast includes the semantic contrast as a proper part. In a previous modeling study, Van Maanen et al. (2009) did not predict differences in the locus of the semantic and Stroop-like effects in PWI, and treated congruent and neutral trials as instances of a single control condition. Moreover, in previous electrophysiological studies (Piai, Roelofs, Jensen, et al., 2014; Roelofs et al., 2016), the Stroop-like effect and the semantic effect were both reflected in the N400: There was a difference in N400 amplitude between incongruent and neutral trials (the semantic effect), and an even larger difference between congruent and incongruent trials (the Stroop-like effect). Thus, if the functional locus of the semantic effect is different between post-congruent and post-incongruent trials, as Van Maanen and Van Rijn claim, then this should also hold for the Stroop-like effect. Therefore, the Stroop-like effect should have been present much earlier in the ERPs at post-congruent trials (i.e., before 200 msec

post picture-onset) than at post-incongruent trials (after 200 msec). In contrast, we observed that the Stroop-like effect was present in the N400 and had a similar onset for post-congruent and post-incongruent trials. Thus, our findings contradict the claim of Van Maanen and Van Rijn that the locus of the PWI effect differs between post-congruent and post-incongruent trials.

Van Maanen and Van Rijn (2010) suggested that they obtained a Gratton effect at the long SOA without reporting appropriate statistics. However, in a re-analysis of their data, we failed to obtain a Gratton effect at the long SOA for the semantic effect. Current trial and previous trial did not interact [$F(1,21) = .64, p = .43$], different from what is suggested by Van Maanen and Van Rijn (2010, p. 172). The Gratton effect was not present for the Stroop-like effect either [$F(1,21) = .1, p = .76$]. These inconsistencies make the results of Van Maanen and Van Rijn problematic.

Since it has remained unclear how robust the findings of Van Maanen and Van Rijn (2010) are, we conducted additional analyses on the dual-task performance data collected by Piai, Roelofs, and Schriefers (2014, Experiment 2). The dual-task procedure used by Piai et al. was highly comparable to that of Van Maanen and Van Rijn. On every trial, a tone-classification stimulus was combined with a PWI stimulus in one of three conditions (i.e., congruent, incongruent, or neutral) either at zero-SOA (short-SOA condition) or at an SOA of 500 msec (long-SOA condition). We restricted our analysis to the semantic and Stroop-like effects in the critical short-SOA condition. For the semantic effect, a 2 × 2 ANOVA with factors current trial (neutral, incongruent) and previous trial (congruent, incongruent) showed a main effect of current trial [$F(1,20) = 16.0, p < .001$], but no main effect of previous trial [$F(1,20) = .14, p = .72$] or interaction of current trial and previous trial [$F(1,20) = .03, p = .87$]. Similarly, for the Stroop-like effect, only a main effect of current trial [$F(1,20) = 13.61, p = .002$] was present, but no main effect of previous trial [$F(1,20) = 2.03, p = .17$] or interaction of current trial and previous trial [$F(1,20) = 2.19, p = .16$]. The absence of an interaction of current trial and previous trial indicates that the magnitude of the semantic and the Stroop-like effects did not differ between post-congruent and post-incongruent trials at the short SOA, unlike what Van Maanen and Van Rijn observed for the semantic effect. Thus, in line with our present electrophysiological findings, the dual-task data of Piai et al. do not provide evidence for a shift in the locus of the PWI effect depending on the previous trial type, which challenges the claim by Van Maanen and Van Rijn.

### 4.2. Electrophysiological manifestations of the Gratton effect

Besides presenting evidence against the idea of a shift in the locus of the Stroop-like effect in PWI, our study is the first to present evidence for a Gratton effect in the electrophysiological patterns of brain activity (i.e., the N400 component and theta power) using overt spoken responses. Several previous studies failed to observe control adjustment effects in the

N400 amplitude using various adaptations of the Stroop task with manual or spoken responses (Donohue et al., 2012; Larson et al., 2009, 2012). Until now, only West et al. (2012) reported a Gratton-like modulation of the Stroop-like effect in a counting Stroop task with manual responses. Given that we did observe the Gratton effect using overt spoken responses, it is important to see in what way our study was different from previous EEG investigations in the field. First, one clear difference between PWI and Stroop is the stimulus set size. We used a relatively large stimulus set consisting of 39 pictures, while the classical Stroop task employs three or four colors. The large stimulus set is an obvious asset of PWI, because it prevents multiple repetitions of the same stimuli throughout the experiment, which has been shown to potentially invalidate top-down control interpretations of the Gratton effect (Hommel et al., 2004; cf. Larson et al., 2016; Mayr et al., 2003). Second, we used a larger overall number of trials in the experiment than some previous studies. This increased our statistical power and yielded a high signal-to-noise ratio (for a discussion of statistical power issues in control adjustments effects in Stroop-like tasks, see review Larson et al., 2014).

Interestingly, the way the Gratton effect manifested itself in our study was different for the RTs and the ERPs. In the RTs, the interaction between current trial type and previous trial type was driven by the current-congruent trials. RTs on congruent-after-incongruent trials were longer than RTs on congruent-after-congruent trials, while RTs on incongruent-after-congruent and incongruent-after-incongruent trials did not differ. However, the interaction in the N400 amplitude was driven by a stronger negative deflection on incongruent-after-congruent trials than on incongruent-after-incongruent trials, while the N400 amplitude on congruent-after-congruent and congruent-after-incongruent trials did not differ. As we discussed in the beginning of this section with respect to discrepancy between RT and error rates patterns, the Gratton effect may take different forms of interaction of current trial and previous trial (Lamers & Roelofs, 2011). The difference in the way the Gratton effect is present in the RTs and ERPs in the present study suggests that an influence of response caution was present in the RTs, but not, or differently, in the ERPs. Greater caution after incongruent than after congruent trials also explains why error rates were reduced (see Fig. 2). Future ERP studies may examine this account further.

To our knowledge, the present study is the first to directly investigate effects of control adjustments on midfrontal theta power in speech production. Midfrontal theta has been shown to originate from the medial frontal cortex including the ACC and it has been related to cognitive control, task switching, error detection, and feedback processing outside the language domain (Debener et al., 2005; Hanslmayr et al., 2008; Luu, Tucker, & Makeig, 2004). Until now only one study reported theta power effects in PWI (Piai, Roelofs, Jensen, et al., 2014). Piai et al. used magnetoencephalography (MEG). Therefore, a direct comparison of scalp-level topographies of the Stroop-like theta power effect observed in their study and

in the present study is problematic. However, the topography observed in the current study is compatible with the left superior prefrontal source (presumably including portion of the ACC) of the theta activity estimated from the MEG data by Piai et al. Our findings regarding control adjustments in speaking revealed by midfrontal theta-band effects are also in line with previous studies investigating control adjustments in manual Stroop and Eriksen flanker tasks (Cavanagh et al., 2009; Hanslmayr et al., 2008; Nigbur et al., 2011).

It remains unclear to what extent our present N400 and theta power results represent the same or distinct brain activity. A comparison of the TFR of the ERP data and the standard single-trial TFR showed that the latter largely reflects phase-locked theta dynamics. Moreover, inter-trial coherence (ITC) analysis, which quantifies across trials the consistency of the frequency-specific phase, suggested a midfrontal theta effect over which the ITC values were increased on incongruent trials compared to congruent trials relative to other sites. However, in neither analysis did we observe a Gratton effect. A presence of the Gratton effect in the theta portion of the ERP and the ITC values would have suggested that our theta power and ERP effects are reflecting the same brain dynamics. However, given the absence of this effect, the current dataset does not allow to make claims regarding a same or distinct origin of the N400 and theta oscillations dynamics. Based on previous work in the field, it would be plausible that the theta activity observed in the current study was generated by the prefrontal cortex including ACC (e.g., Hanslmayr et al., 2008; Piai, Roelofs, Jensen, et al., 2014). However the source of the N400 activity in PWI remains unclear, being either ACC (like in other Stroop-like tasks; Hanslmayr et al., 2008) or left middle temporal gyrus (Piai, Roelofs, Jensen, et al., 2014).

## 5. Conclusions

The present study provided evidence on control adjustments in speaking using the standard PWI task. The Gratton effect manifested itself in RTs, error rates, amplitude of the N400 component, and midfrontal theta power. The onset of the ERP effect at 350–400 msec post stimulus-onset (i.e., in the N400) suggests that the Stroop-like effect occurred at the word planning stage regardless of the previous trial type. This finding contradicts the account of Van Maanen and Van Rijn (2010) that posits that the functional locus of the PWI effect shifts depending on whether the previous trial is congruent or incongruent. Moreover, we observed a Gratton effect in the midfrontal theta power, which is in line with the research on the role of theta-band oscillations in cognitive control, until now almost exclusively reported for manual tasks. To conclude, our results provide evidence that the PWI effect

arises at the word planning stage following both congruent and incongruent trials, while the amount of top-down control changes depending on the previous trial type.

## Acknowledgements

## Appendix A. Stimulus list.

| Category | Picture name | Incongruent distractor |
| --- | --- | --- |
| Kitchenware | pan (pan) | bord |
| | bord (plate) | fles |
| | fles (bottle) | pan |
| Clothing | trui (sweater) | broek |
| | broek (trousers) | jas |
| | jas (jacket) | trui |
| Transportation | bus (bus) | trein |
| | trein (train) | vliegtuig |
| | vliegtuig (airplane) | bus |
| Buildings | toren (tower) | fabriek |
| | fabriek (factory) | kasteel |
| | kasteel (castle) | toren |
| Body parts | neus (nose) | voet |
| | voet (foot) | oog |
| | oog (eye) | neus |
| Fruit | peer (pear) | banaan |
| | banaan (banana) | appel |
| | appel (apple) | peer |
| Food | worst (sausage) | kaas |
| | kaas (cheese) | brood |
| | brood (bread) | worst |
| Birds | kalkoen (turkey) | duif |
| | duif (pigeon) | haan |
| | haan (roster) | kalkoen |
| Cutlery | lepel (spoon) | vork |
| | vork (fork) | mes |
| | mes (knife) | lepel |
| Tools | hamer (hammer) | tang |
| | tang (pliers) | zaag |
| | zaag (saw) | hamer |
| Building parts | dak (roof) | trap |
| | trap (stairs) | muur |
| | muur (wall) | dak |
| Bags | rugzak (backpack) | koffer |
| | koffer (suitcase) | tas |
| | tas (bag) | rugzak |
| Furniture | kast (cupboard) | stoel |
| | stoel (chair) | tafel |
| | tafel (table) | kast |

## Appendix B. Median number of trials averaged for individual ERPs.

The shaded area marks the 25th and 75th percentiles.
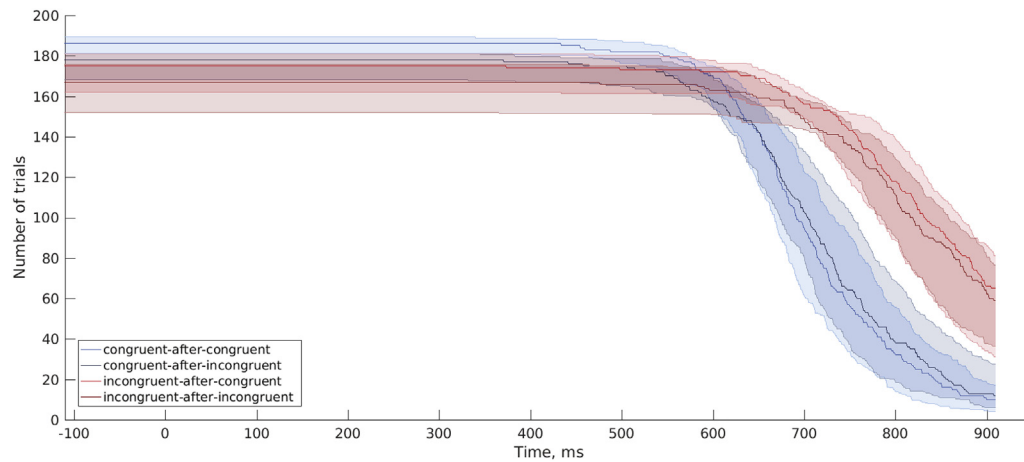


## Appendix C
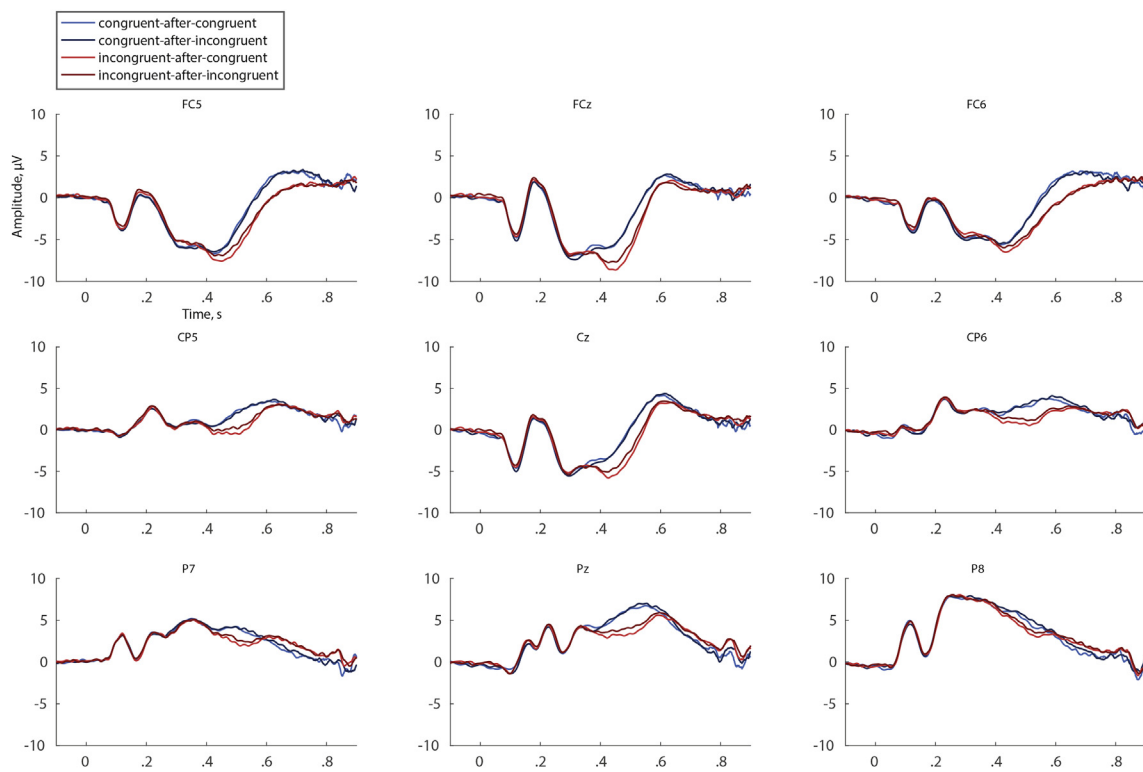
# REFERENCES

Aarts, E., & Roelofs, A. (2011). Attentional control in anterior cingulate cortex based on probabilistic cueing. *Journal of Cognitive Neuroscience, 23*, 716–727.

Aarts, E., Roelofs, A., & Van Turennout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *The Journal of Neuroscience, 28*, 4671–4678.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.

Botvinick, M. M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179–181.

Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., et al. (2000). Parsing executive processes: Strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 1944–1948.

Cavanagh, J. F., Cohen, M. X., & Allen, J. J. (2009). Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. *The Journal of Neuroscience, 29*, 98–105.

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology, 49*, 220–238.

Clayson, P. E., & Larson, M. J. (2011). Conflict adaptation and sequential trial effects: Support for the conflict monitoring theory. *Neuropsychologia, 49*, 1953–1961.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *The Journal of Neuroscience, 25*(50), 11730–11737.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is a Stroop effect. *Psychonomic Bulletin & Review, 14*, 717–722.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology, 1*, 1–10.

Donohue, S. E., Liotti, M., Perez, R., III, & Woldorff, M. G. (2012). Is conflict monitoring supramodal? Spatiotemporal dynamics of cognitive control processes in an auditory Stroop task. *Cognitive, Affective, & Behavioral Neuroscience, 12*, 1–15.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 380–390.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience, 8*, 1784–1790.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*, 152–170.

Freund, M., Gordon, B., & Nozari, N. (2016). Conflict-based regulation of control in language production. In *Proceedings of the 38th annual conference of the cognitive science society*.

Glaser, W. R., & Düngelhoff, F. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640–654.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*, 480–506.

Hanslmayr, S., Pastötter, B., Bäuml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience, 20*, 215–225.

Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research, 68*, 1–17.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology, 2*, Article 255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*, 1023–1026.

Kopp, B., Rist, F., & Mattler, U. W. E. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology, 33*, 282–294.

Lamers, M. J., & Roelofs, A. (2011). Attentional control adjustments in Eriksen and Stroop task performance can be independent of response conflict. *The Quarterly Journal of Experimental Psychology, 64*, 1056–1081.

Larson, M. J., Clawson, A., Clayson, P. E., & South, M. (2012). Cognitive control and conflict adaptation similarities in children and adults. *Developmental Neuropsychology, 37*, 343–357.

Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology, 93*, 283–297.

Larson, M. J., Clayson, P. E., Kirwan, C. B., & Weissman, D. H. (2016). Event-related potential indices of congruency sequence effects without feature integration or contingency learning confounds. *Psychophysiology, 53*, 814–822.

Larson, M. J., Kaufman, D. A., & Perlstein, W. M. (2009). Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia, 47*, 663–670.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38.

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia, 38*, 701–711.

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology, 115*(8), 1821–1835.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*, 177–190.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience, 6*, 450–452.

Nigbur, R., Ivanova, G., & Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical Neurophysiology, 122*, 2185–2194.

Notebaert, W., Gevers, W., Verbruggen, F., & Liefooghe, B. (2006). Top-down and bottom-up sequential modulations of congruency effects. *Psychonomic Bulletin & Review, 13*, 112–117.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 1–9.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 358–377.

Pastötter, B., Dreisbach, G., & Bäuml, K. H. T. (2013). Dynamic adjustments of cognitive control: Oscillatory correlates of the conflict adaptation effect. *Journal of Cognitive Neuroscience, 25*, 2167–2178.

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J. M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLoS One, 9*, e88674.

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 147–165.

Piai, V., Roelofs, A., & Van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research, 1450*, 87–101.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110*, 88–125.

Roelofs, A. (2014). Tracking eye movements to localize Stroop interference in naming: Word planning versus articulatory buffering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1332–1347.

Roelofs, A., Piai, V., Garrido Rodriguez, G., & Chwilla, D. J. (2016). Electrophysiology of cross-language interference and facilitation in picture naming. *Cortex, 76*, 1–16.

Schnur, T. T., & Martin, R. (2012). Semantic picture–word interference is a postperceptual effect. *Psychonomic Bulletin & Review, 19*, 301–308.

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M. (2016). Using brain potentials to functionally localise Stroop-like effects in colour and picture naming: Perceptual encoding versus word planning. *PLoS One, 11*, e0161052.

Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: It's not just priming. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 467–472.

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*, 584–589.

Van Maanen, L., & Van Rijn, H. (2010). The locus of the Gratton effect in picture–word interference. *Topics in Cognitive Science, 2*, 168–180.

Van Maanen, L., Van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review, 16*, 987–999.

West, R., & Bailey, K. (2012). ERP correlates of dual mechanisms of control in the counting Stroop task. *Psychophysiology, 49*, 1309–1318.

West, R., Bailey, K., Tiernan, B. N., Boonsuk, W., & Gilbert, S. (2012). The temporal dynamics of medial and lateral frontal neural activity related to proactive cognitive control. *Neuropsychologia, 50*, 3450–3460.