



Is switching more costly in cued than voluntary language switching? Evidence from behaviour and electrophysiology

Nora Kennis^{1,2} , Xiaochen Y. Zheng¹ , Angela de Bruin³ and Vitória Piai^{1,4}

Research Article

Cite this article: Kennis, N., Zheng, X.Y., de Bruin, A. and Piai, V. (2024). Is switching more costly in cued than voluntary language switching? Evidence from behaviour and electrophysiology. *Bilingualism: Language and Cognition*, 1–14
<https://doi.org/10.1017/S1366728924000452>

Received: 13 June 2023
Revised: 20 May 2024
Accepted: 20 May 2024

Keywords:

language switching; language control; electrophysiology; language proficiency; inhibition

Corresponding author:

Nora Kennis;
Email: nora.kennis@ed.ac.uk

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

¹Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; ²Department of Psychology, University of Edinburgh, Edinburgh, UK; ³Department of Psychology, University of York, York, UK and ⁴Department of Medical Psychology, Radboudumc, Donders Centre for Medical Neuroscience, Nijmegen, The Netherlands

Abstract

Multilingual language control is commonly investigated using picture-naming paradigms with explicit instructions when to switch between languages. In daily life, language switching also occurs without external cues. Cued language-switching tasks usually show a switch cost (i.e., slower responses on switch than non-switch trials). Findings of switch costs in response times are mixed for voluntary language switching. This pre-registered study uses a bilingual picture-naming paradigm to compare voluntary and cued language switching in 25 highly proficient Dutch-English bilinguals using EEG. We analysed the N2 ERP component and midfrontal theta oscillations, two common electrophysiological markers of cognitive control in task and language switching. We observed significantly smaller behavioural switch costs in the voluntary task. This suggests that voluntary language switching is less effortful than switching based on external cues. However, we found no electrophysiological switch effects in either task. We discuss factors which may contribute to the inconsistency between behavioural and electrophysiological findings.

Highlights

- EEG methodologies could tell us more about language control in bilingual contexts.
- We compare voluntary and cued switching using RTs, ERPs and oscillations.
- Voluntary language switching is behaviourally less effortful than cued switching.
- There were no language-switching effects in the EEG measures.
- The link between behavioural and neural switch cost might be weaker than expected.

1. Introduction

Bilinguals can fluently switch between their languages in everyday circumstances. These language switches might require control, so speakers can select the appropriate target language at the right time and avoid interference from other languages. Some switches are imposed by the interactive context, for example, when a new person who does not speak the current language joins a conversation. Laboratory studies have examined these processes using cued picture-naming paradigms in which participants are instructed when to switch between languages (Costa & Santesteban, 2004; Meuter & Allport, 1999). In daily life, however, bilinguals can also switch between their languages freely when interlocutors share multiple languages, a process that is rarely studied using brain-based measures. The present study uses electrophysiology to investigate switch effects in voluntary and cued language switching by examining markers of language control.

1.1 Language-switching paradigms

Cued tasks typically used to study language switching in laboratory settings require participants to name pictures or digits in both of their languages, with the language determined by a cue (often a colour or a flag). These studies usually show that speakers are slower to respond on switch trials (the language of the current trial is different from the previous trial) than on repeat trials (the language of two consecutive trials is the same). This phenomenon is known as a switch cost (Meuter & Allport, 1999) and has been observed across a range of bilinguals, both those whose proficiency in their first and second languages is approximately matched (balanced bilinguals, e.g., Costa & Santesteban, 2004), and those who are more proficient in one of their languages (unbalanced bilinguals, e.g., Li et al., 2021; Philipp et al., 2007; Verhoef et al., 2009).

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



Cued language-switching tasks reflect real-life contexts in which bilinguals switch languages on demand, such as between interlocutors: a speaker may use one language with a friend at school and another with a teacher. A switching type many studies leave unexplored, however, is voluntary language switching. When a bilingual is surrounded by others who share one or more of their languages, they may decide to switch between languages as they see fit or as is easiest. Multiple recent behavioural studies have investigated voluntary language switching in comparison to cued switching. Their participants perform one traditional cued picture-naming task and one task with free language choice for each picture. These studies typically find faster voluntary than cued overall response times (e.g., de Bruin *et al.*, 2018; Jevtović *et al.*, 2020). Looking at switching in particular, results are more mixed: some studies find that voluntary switching can still lead to a switch cost in response times comparable to cued switching (e.g., de Bruin *et al.*, 2018, Experiment 1; Gollan *et al.*, 2014). Others show that the behavioural switch cost for voluntary switching is smaller than in cued switching (Gollan *et al.*, 2014, Experiment 2; Jevtović *et al.*, 2020) or that the switch cost could even be absent in voluntary switching (Blanco-Elorrieta & Pyllkänen, 2017; Kleinman & Gollan, 2016; Zhu *et al.*, 2022).

Many of the studies that report no voluntary switch cost are designed for language switching to be solely based on lexical accessibility, with switching to another language being easier because the word is more readily available in that language. This effect is either created artificially by asking participants to choose whichever language is easiest at the first occurrence of a stimulus item and then to continue naming that item in that language at every repetition (Kleinman & Gollan, 2016) or naturally by using items almost uniquely associated with one of the languages (Zhu *et al.*, 2022). Consistency of language per item and reliance on lexical access might increase efficiency and reduce or eliminate the switch cost. In contrast, language-switching studies without these specific instructions often do find a cost. This includes corpus-based analyses assessing switching in conversations (Fricke *et al.*, 2016), suggesting switch costs are not an artefact of laboratory-based studies. Indeed, there can be several reasons why bilinguals do or do not switch languages beyond lexical access (e.g., Fricke & Kootstra, 2016), including preference for using a specific default language or alignment with an interlocutor.

As for the cognitive processes behind the various language switch costs, most studies consider the involvement of cognitive control, which refers to the general mechanisms that “regulate thought and actions in accordance with internally represented behavioural goals” (Braver, 2012, p. 106). For example, the Inhibitory Control model (Green, 1998) posits that words in the non-target language are inhibited to facilitate the use of the target language. The switch cost then reflects at least two aspects of language control: inhibiting the new non-target language and activating the target language that was inhibited in the previous trial. A possible explanation for the difference between cued and voluntary language switching is offered by the Adaptive Control Hypothesis (ACH; Green & Abutalebi, 2013): These different switching contexts vary in the type and level of language control required. Bilingual contexts in which language switches are determined by outside cues (such as interlocutor identity) require processes such as goal monitoring, cue detection and response inhibition. These ensure the speaker maintains a constant awareness of the appropriate language and can suppress responses in the non-target language. Conversely, in contexts with dense code-switching (more akin to voluntary switching), less control might be needed to monitor cues and inhibit responses. In these environments,

bilinguals might use words that come to mind fastest, regardless of the language (e.g., de Bruin *et al.*, 2018).

1.2 Electrophysiology of language switching and cognitive control

The role of cognitive control in language switching is echoed in the electrophysiological markers that have been established for language switching for both event-related potentials (ERPs) and time-frequency representations (TFRs). The N2 ERP component is a negative peak around 200–350 ms post-stimulus onset, which is associated with various types of top-down control in speech production literature (Schmitt *et al.*, 2000; Trewartha & Phillips, 2013). More generally, the N2 component has been found in relation to response inhibition (e.g., Jodo & Kayama, 1992) and conflict monitoring (e.g., Donkers & Van Boxtel, 2004). A number of language-switching studies have also found that the frontal or posterior negative peak had a higher amplitude in language switch trials than repeat trials during the N2 time window (e.g., Declerck *et al.*, 2021; Jackson *et al.*, 2001; Verhoef *et al.*, 2010; Zheng *et al.*, 2020). A prevalent interpretation of this N2 in language-switching tasks has been that it plays a role in the active inhibition of the speaker’s first language (L1) to gain access to their second language (L2) (e.g., Jackson *et al.*, 2001), or in the overcoming of inhibition of a languages (e.g., Verhoef *et al.*, 2010).

Research on neuronal oscillations has furthermore found higher midline frontal theta oscillations to be associated with increased executive control (e.g., Cavanagh & Frank, 2014; Cohen & Donner, 2013; Cooper *et al.*, 2019). For example, a recent study into non-linguistic tasks that require various types of cognitive control found higher midfrontal theta power for trials that require more control (Eisma *et al.*, 2021). Switch effects in theta-band power have also been found in task switching (Cooper *et al.*, 2017), with higher power on switch than repeat trials. Research into language switching and neuronal oscillations is scarcer, but midfrontal theta power has been shown to increase after speakers have selected the incorrect (i.e., non-target) language versus correct responses (Piai & Zheng, 2019), signalling increased recruitment of top-down control (c.f., Cavanagh *et al.*, 2012; Luu *et al.*, 2004). Similarly, a number of monolingual studies note increased theta power in response to semantically related distractors in picture-word interference tasks (Krott *et al.*, 2019; Piai, Roelofs, Jensen, *et al.*, 2014; Shitova *et al.*, 2017). This effect could reflect the increased cognitive and language control demands resulting from the conflict between the target and the distractor. More recently, Cui *et al.* (2024) showed a midfrontal theta power increase in language switch trials compared to repeat trials (i.e., switch cost).

Although there are only a small number of studies, the link between midfrontal theta oscillations and executive control seems to extend to language control, with theta power increasing when more language control is required. It also seems clear that the N2 ERP component is linked to the cognitive control required for language switching. Still, electrophysiological studies on the comparison between cued and voluntary switching remain rare. One example is a recent MEG study that located the cued switch effect in the prefrontal cortex, which is commonly engaged in effortful task switching. Where cued switching engaged the prefrontal cortex, voluntary switching did not (Blanco-Elorrieta & Pyllkänen, 2017). Further electrophysiological studies could help us gain more insights into the neurocognitive processes behind voluntary switching. What happens to the brain when bilinguals are allowed

to switch whenever they prefer? Can we still observe the switching cost found in cued switching behaviourally and electrophysiologically?

1.3 Current study

The current pre-registered study aimed to investigate switch effects in voluntary compared to cued language switching as well as their electrophysiological manifestation. The electrophysiology of voluntary switching has remained largely unexplored but could provide new insights into the neural processes of language control in multilinguals. We collected behavioural and EEG data from adult Dutch-English late bilinguals performing two bilingual picture-naming tasks. In one task, participants were cued on the language for each trial by the picture's background colour, while in the other, they were free to decide which language to use for each trial. We selected participants with high proficiency and frequent use of English in daily life to increase the expected rate of English use in the voluntary-switching task. Our investigation of the EEG data focussed on the N2 ERP component and midline frontal theta oscillations.

Behaviourally, we expected responses to be slower on switch than repeat trials (e.g., de Bruin et al., 2018). In terms of electrophysiological data, we predicted an N2 switch effect within each task based on previous studies (Jackson et al., 2001; Verhoef et al., 2010; Zheng et al., 2020), as well as a midfrontal theta switch effect based on experiments on task switching and language switching (Cooper et al., 2017; Cui et al., 2024).

In terms of overall task effects, cued switching could be more demanding according to the ACH (Green & Abutalebi, 2013), and earlier behavioural studies show a significant task effect of longer response times (RTs) in cued tasks. We therefore expected an N2 task effect, with higher N2 amplitudes for trials in the cued compared to the voluntary task. Similarly, we expected higher midfrontal theta power in cued compared to voluntary trials considering non-linguistic trials that require more cognitive control have been associated with increased midfrontal theta oscillations (Eisma et al., 2021).

Based on previous literature, an interaction between switch effects and task effects is plausible: the voluntary switch cost could be smaller than cued costs (Blanco-Elorrieta & Pylkkänen, 2017; Jevtović et al., 2020) or the same (de Bruin et al., 2018; Gollan et al., 2014). Electrophysiological switch costs may also differ between the task contexts. If cued switching recruits more control behaviourally than voluntary switching, we expect the cued task to have a greater electrophysiological switch cost than the voluntary task. Alternatively, if voluntary switching is equally effortful as cued switching, as evidenced by the RTs, electrophysiological switch costs might not differ.

2. Methods

The pre-registration for this project can be found at <https://osf.io/6tdhe>.

2.1 Participants

Twenty-five Dutch-English late bilinguals took part in the experiment and were compensated with study credits or vouchers. This number of participants was determined as follows: most EEG language-switching studies recruit between 20 and 30 participants for within-subject designs. Recruiting 25 participants in a within-

subject design allowed us to detect effects with an effect size of at least $d = .59$ with a power of .8, based on a calculation for a two-tailed paired t -test using the R pwr package (version 1.3–0; Champely, 2020). Admittedly, this is not a perfect calculation, as this type of study comparing cued and voluntary switching had not been done before. This study could serve as a first step to determine the effect sizes for the EEG on which future studies can build. Furthermore, participant recruitment was limited to this number due to budget and time constraints. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The study was approved by the local ethics committee (Faculty Ethics Committee, Radboud University, ECSW 2017-3001-455). All participants provided written informed consent.

All participants were native Dutch speakers who spoke English regularly in their daily life because of their study programme or job and rated their English to be “very good” (the highest level on a five-point Likert scale). Participants (6 male, 19 female) were 22.4 years old on average ($SD = 2.6$, range 18–27). Table 1 contains a summary of their self-rated language proficiency and usage, as well as their average LexTALE score to indicate L2 proficiency, which is “advanced/proficient” (Lemhöfer & Broersma, 2012). All participants were right-handed, had a normal or corrected-to-normal vision and had no known neurological, speaking or hearing impairments.

2.2 Materials

We used 40 colour pictures representing noncognate pairs in Dutch and English, divided into two subsets of 25 images (A and B), one for each task. The pictures were taken from the

Table 1. Summary of language background questionnaire

	Dutch			English		
	Mean	SD	Range	Mean	SD	Range
Age of acquisition (years)	0	0	0	9.8	1.9	4–12
LexTALE score (0–100%)				87	10.6	56–100
Self-rated proficiency ^a						
Speaking	9.4	0.8	8–10	8.0	0.8	7–10
Listening	9.6	0.8	8–10	9.1	0.8	7–10
Reading	9.5	0.8	8–10	9.0	1.0	7–10
Writing	8.9	1.2	7–10	7.9	1.1	7–10
Frequency of use ^a						
Speaking	7.6	1.5	4–10	7.6	1.6	5–10
Listening	7.8	2.3	3–10	9.0	0.9	7–10
Reading	7.0	2.7	1–10	8.7	1.2	5–10
Writing	7.9	2.1	3–10	7.2	1.8	3–10
Frequency of switching ^a						
Within a day	7.5		1.4	4–10		
Within a conversation	4.9		2.1	2–10		
Within a sentence	4.6		2.5	2–10		

^aSelf-ratings were provided on a scale of 1 (very bad/never) to 10 (very good/all the time).

MultiPic database (Duñabeitia *et al.*, 2018), a selection was made in which naming agreement in Dutch and English was over 70% (naming agreement, i.e., percentage of participants using the most frequent name: $M_{NL} = 94.97$, $SD_{NL} = 7.00$, $M_{EN} = 92.91$, $SD_{EN} = 8.83$) and labels were appropriately high frequency (Zipf frequency: $M_{NL} = 4.26$, $SD_{NL} = 0.58$, $M_{EN} = 4.38$, $SD_{EN} = 0.50$; SUBTLEX-UK database: van Heuven *et al.*, 2014; SUBTLEX-NL: Keuleers & Brysbaert, 2010). The names of the pictures were matched as closely as possible between English and Dutch and between sets A and B for frequency ($t_{A-B}(78) = -0.54$, $p_{A-B} = .593$; $t_{NL-EN}(78) = -1.00$, $p_{NL-EN} = .320$), number of syllables ($t_{A-B}(78) = 0.17$, $p_{A-B} = .861$; $t_{NL-EN}(78) = -0.52$, $p_{NL-EN} = .601$) and number of phonemes ($t_{A-B}(78) = -0.13$, $p_{A-B} = .900$; $t_{NL-EN}(78) = -0.25$, $p_{NL-EN} = .802$). We also matched the words for onset phoneme category, meaning there was an equal number of words starting with fricatives, vowels, plosives, and so forth between English and Dutch and sets A and B (see Appendix A for a full list of stimuli). A separate set of images depicting 10 more non-cognate pairs with high naming agreement and high frequency was used for training.

2.3 Design

The experiment had two types of trials: repeat trials (the language of the trial matches the previous trial) and switch trials (the language of the trial differs from the previous trial), both present in the voluntary and cued switching tasks. This led to a 2×2 within-subject design with the predictors of interest task type (cued/voluntary) and trial type (switch/repeat). Language was not included in the main analysis as there would be too few trials in each condition for a reliable analysis.

Both tasks consisted of three blocks of 80 trials each, leading to a total of 240 trials per task. In the cued task, this was split into 80 switch and 160 repeat trials, with half of the trials being Dutch and half being English. This ratio of switch versus repeat trials was chosen based on rates reported in previous voluntary-switching studies with unbalanced bilinguals (e.g., de Bruin & Xu, 2022) and confirmed by way of a small pilot study of four participants who switched on 40% of voluntary trials on average.

The order of stimuli presentation was pseudo-randomised using the programme MIX (Van Casteren & Davis, 2006) with a number of restrictions. In both the voluntary and the cued task, the randomisation constraints were that (1) two subsequent trials could not have the same semantic category or the same onset phoneme (in either language) and (2) there were at least five trials between occurrences of the same image. Additional constraints for the cued task were that (3) the maximum number of consecutive switches was two, (4) the maximum number of consecutive trials in the same language was five and (5) there were at least ten trials between the occurrence of the same image in the same language.

2.4 Procedure

Participants started the experimental session with a stimulus familiarisation task to minimise errors or switching behaviours caused by speakers not knowing the name for an image. This was followed by the two picture-naming tasks (the order of the cued and voluntary tasks was counterbalanced between participants), during which EEG was recorded. Finally, they completed a short English vocabulary test (LexTALE, Lemhöfer & Broersma, 2012) and a language background questionnaire. Full testing sessions (including EEG preparation) lasted 2–2.5 hours. The experimenter made sure to

switch between English and Dutch regularly while preparing the participants for the experiment to establish a multilingual environment.

During stimulus familiarisation, participants first named all 50 pictures in Dutch, then again in English. Items of which they did not know the English name were repeated one more time. The correct name for the item would appear after the participant's response. In the case of an error, participants repeated the correct name. Next came the two experimental tasks, starting with either the voluntary or the cued picture-naming task. Participants were asked to name pictures in English and Dutch in both tasks. All tasks were displayed on a computer monitor (Benq XL2420Z, 24-inch screen with a resolution of 1920 × 1080 pixels) with a grey background using Presentation software (Version 20.3; Neurobehavioural Systems Inc, 2021). Stimulus images were scaled to 300 × 300 pixels, surrounded by a 100-pixel wide-coloured frame (background). Trials started with a fixation cross for 1000 ms, after which the picture (with coloured border, see below) was presented. Participants' responses were registered using a voicekey (with a Shure SM-57 microphone). After each response or after the time limit of 3000 ms had passed, the stimulus remained on the screen for another 550 ms, followed by a jittered blank screen (between 250 and 500 ms).

The image set/task-type combination was counterbalanced between participants so half of the participants performed the cued task on image set A and the voluntary task on set B, and the other half of participants received the opposite combination. All instructions for both tasks were displayed on the screen in Dutch before the practice block, then in English after the practice block. Short reminders of the instructions displayed between blocks were interchangeably in Dutch and English. We asked participants to name the pictures as quickly and accurately as possible, to not correct themselves if they had made a mistake, and to avoid saying anything but the picture names (like “uhm”) so the voicekey would not be erroneously triggered.

In the cued task, the colours of the image background cued the language of the trial. The use of unnatural cues like colours in language-switching tasks has been previously criticised, and several papers have attempted to use more natural cues such as faces of interlocutors (Blanco-Elorrieta & Pykkänen, 2017; Zhu *et al.*, 2022). While we considered this option, we anticipated that displaying the cues either before or next to the stimulus would affect the nature or quality of our EEG data too strongly to remain usable. Therefore, blue/green backgrounds cued one language and orange/red backgrounds cued another. Two colours were used per language and the frame colour switched in every trial to avoid confounds of the effects of cue switches in the stimulus and language switches in the responses (c.f., Heikoop *et al.*, 2016; Mayr & Kliegl, 2003). Which language was assigned to which cue combination was counterbalanced across participants. They performed a practice block to ensure cue familiarity, which ran for at least 40 trials and ended once 90% total accuracy was reached after that. Feedback of the correct word and language was provided up to practice trial 20. After each experimental block of 80 trials, participants could take a break.

In the voluntary task, the language of the trial did not need to be cued using image backgrounds. Still, to match the cued task as closely as possible, the voluntary task had interchanging yellow and purple backgrounds. Participants first performed 20 practice trials to ensure the task was understood (longer training was not required). Feedback on the two correct responses (one Dutch, one English) was provided up to practice trial 10. The instructions

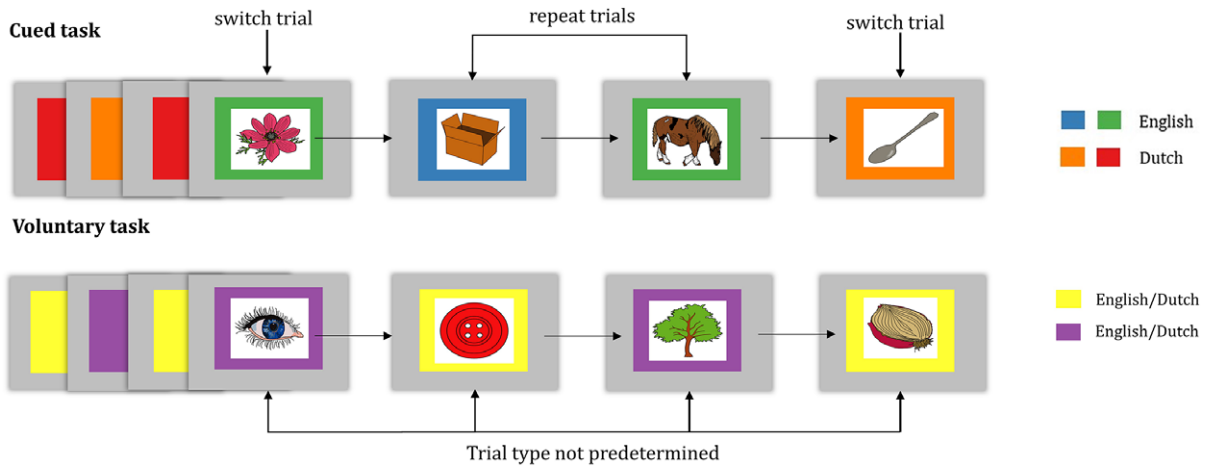


Figure 1. Experimental paradigm for the cued and voluntary tasks. Stimuli were counterbalanced between tasks. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/YWAVN>.

for this task said “please name these pictures in English or in Dutch. You are free to pick English or Dutch when naming each item. Make sure to use both languages during the task.” It was never the case that participants only responded in one language or switched languages at every trial. They were allowed to take a break after each of the three blocks of 80 trials.

Figure 1 shows a schematic diagram of the experimental paradigm, and the distinction between the voluntary and cued tasks.

2.5 EEG data acquisition

We recorded EEG from 60 active scalp electrodes placed in an elastic cap according to the international 10–20 system using the 64-channel ActiCAP system (Brainproducts), amplified with BrainAmps DC amplifiers (500 Hz sampling, 0.016–125 Hz bandpass filter). All electrodes were referenced online to the left mastoid electrode and re-referenced offline to the averaged mastoids. The electro-oculogram (EOG) was recorded horizontally using two electrodes placed next to the left and the right eye on the temples and vertically from one electrode placed below the left eye. Impedances for electrodes were kept below 20 k Ω during the experiment.

2.6 EEG pre-processing

We performed data pre-processing and analysis using MATLAB (R2020b, The Math Works, Inc) and the Fieldtrip toolbox version 20220208 (Oostenveld et al., 2011). Trials were segmented from the continuous EEG signal into trial epochs of 500 ms before to 1000 ms after stimulus onset. The raw signal was baseline-corrected based on the average EEG activity in the 500 ms before stimulus onset. This process of demeaning serves largely the same purpose as a high-pass filter by correcting pre-stimulus drift (see e.g., Tanner et al., 2015; Widmann et al., 2015). After this, we applied a low-pass two-pass Butterworth filter with a 40 Hz cut-off. Data were then re-referenced to the average of the left and right mastoid electrodes.

We first discarded all error trials (non-target responses, no responses and responses with RT > 3 seconds), post-error trials and the first trials of each block, leading to a mean of 454 ($SD = 17$) retained trials per participant. We visually inspected the data and rejected trials with atypical artefacts (such as jumps and drifts). Trials containing blinks that overlapped with stimulus presentation

were not rejected, meaning the dataset may include a very small number of trials with incorrect time-locking. Since stimulus and cue were presented simultaneously, this should not have affected different conditions asymmetrically. Individual channels with overall bad signals were removed from the data as well, at an average of 1.1 channels ($SD = 0.8$) per participant. Artefacts due to blinks and saccades, and on rare occasions at a few channels, were removed using independent component analysis (ICA). After pre-processing, all participants still had more than the exclusion threshold of 30 trials per condition. Finally, we randomly selected an equal number of trials from each of the conditions, based on each participant’s condition with the lowest number of valid trials. This ensures similar signal-to-noise ratios across various conditions as far as the number of trials is concerned.

2.7 Data analysis

2.7.1 Planned analyses: behavioural

We performed behavioural analysis using R (version 4.1.1; R Core Team, 2021). We coded the accuracy of all trials and the language of the voluntary trials during experimental sessions, which was checked after each session in case of uncertainties. The voluntary task’s trials were coded for trial type (switch/repeat) based on preceding trials using a custom R script. Trial type in the cued task was pre-determined by the pseudo-randomised trial order. To match EEG pre-processing, we removed error trials, post-error trials and the first trial of each block from the behavioural analysis. RTs were recorded by a voicekey. For all RTs under 500 ms, we manually checked the speech waveform using Praat (Boersma & Weenink, 2022), as these latencies are less likely to be true responses, based on fastest picture-naming latencies in language-switching experiments starting around 600–700 ms (de Bruin et al., 2018; Zheng et al., 2020). Where the actual speech onset did not align with the voicekey’s detection, we corrected the RT value.

Statistical analyses of the behavioural data were performed with linear mixed-effects models (lme models) using the lme4 package (version 1.1–27.1; Bates et al., 2015) and the afex package (version 1.3–0; Singman et al., 2023). RTs were log-transformed before analysis because the residuals of the lme models were not normally distributed. Both two-level categorical predictors task type (voluntary = -0.5 , cued = $+0.5$) and trial type (repeat = -0.5 , switch = $+0.5$) were sum-to-zero coded to reduce collinearity.

We started with a full model of log-transformed RTs as a function of the fixed effects task type, trial type and their interaction, which included a maximal random effects structure with all random intercepts and slopes for participants and items. We then removed random slopes step-wise (first by-item, then by-participant) as long as the model did not converge. This entailed removing the item slopes that accounted for the lowest variance until the model reached convergence. P-values were computed using the sjPlot package (version 2.8.10; Lüdtke, 2021). Following the significant interaction, the data set was split by task to separately examine the cued and voluntary switch costs. Appendix B contains all lme models used in the final analyses, both planned and exploratory.

2.7.2 Planned analyses: EEG

We computed ERPs by averaging over trials per participant for each of the four conditions, as well as for all cued task trials, all voluntary task trials, all repeat trials, and all switch trials. The ERP data were then baseline-corrected again based on the 500 ms before stimulus onset. The ERP analyses were focussed on the N2 component, defined as the second negative peak time-locked to the stimulus. The relevant time window was determined to be 180–300 ms post-stimulus by inspecting the grand average of all trials blind to task and trial type and identifying the average N2 peak. All channels were used for this analysis because the topography of the N2 component (anterior or posterior) is debated. For each participant, we computed the switch cost by subtracting the ERP amplitude of the repeat trials from the ERP amplitude of the switch trials for both tasks separately.

TFRs of power were computed for the epochs time-locked to stimulus onset at frequencies ranging between 2 and 30 Hz, using a sliding time window of three cycles advanced in steps of 5 ms and 1 Hz. The data in each time window were then multiplied with a Hanning taper, followed by a Fourier transform. We did this per participant for each of the four conditions, as well as for all cued and voluntary task trials, all repeat trials, and all switch trials. Because we were interested in the midfrontal theta effect, we extracted time-resolved power from the following midfrontal channels: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2 and C4. Switch cost in the form of time-resolved power per participant was computed as follows for the cued and voluntary tasks separately: switch cost = (switch condition – repeat condition)/((switch condition + repeat condition)/2). This results in normalised power changes, that is the difference in power between the two conditions normalised by their mean.

Statistical analyses of the data were performed using non-parametric cluster-based permutation tests (Maris & Oostenveld, 2007) to evaluate the effects averaged over trials per participant. The input to the analyses for the ERPs was the ERP amplitudes in the 180–300 ms time window, per participant, per condition. For the oscillatory effects, analyses were done on time-resolved midfrontal theta power per participant, per condition. The time window and frequency range used for this analysis were from stimulus onset up to 700 ms post-stimulus onset and 4–8 Hz (the theta-band).

First, we compared cued switch trials to cued repeat trials (cued switch effect), then voluntary switch to voluntary repeat trials (voluntary switch effect). The general switch effect constituted the comparison of all switch trials to all repeat trials. The fourth comparison tested for an interaction between task and trial type; we compared the switching cost (as computed above for the ERPs and time-frequency data) between the cued and voluntary tasks. For the main task effect, all trials in the cued task were compared to all

voluntary task trials. The permutation test searched the channel-time (-frequency, for oscillatory effects) space for neighbouring time-points and channels (and frequencies) with similar differences across conditions.

For both ERP and time-frequency analyses, channel neighbours were set so that each channel's neighbours were the directly adjacent channels and the channels diagonally across from that channel. We used the Monte Carlo method with 1000 random permutations to calculate the permutation p-value. The false alarm rate due to multiple comparisons was controlled at the alpha level of 0.05.

2.7.3 Exploratory analyses: by-language analysis

To further explore the data, we pre-registered extra analyses as exploratory as we foresaw their potential benefit. While language was not used as a predictor in the planned analyses, many previous behavioural and electrophysiological studies have found differences between L1 and L2 trials in language-switching experiments (e.g., Jackson et al., 2001; Verhoef et al., 2010; Zheng et al., 2020). To compare our RT results more closely to these previous studies, language was added to the lme model (see Appendix B). We decided to forego adding language as a variable to the EEG analyses, as those analyses would be underpowered which would not lead to better comparability to other work.

3. Results

3.1 Planned analyses

3.1.1 Behavioural analysis

Behavioural accuracy showed that participants performed well in all conditions. The average accuracy on the cued task was 96.3% ($SD = 3.5$) and 99.6% ($SD = 0.6$) on the voluntary task. Voluntary-switching frequency was also close to the hypothesised switching frequency of 33% (the fixed switching rate in the cued task): mean switching frequency was 37.4% ($SD = 6.0$), ranging from 23% to 48%. On average, participants spoke Dutch in 46.8% of all trials ($SD = 9.6$) and English in 52.9% of all trials ($SD = 9.6$) in the voluntary task, suggesting that the distribution of languages was comparable for the cued and voluntary tasks.

Figure 2 shows individual mean naming latencies as a function of task type (voluntary/cued) and trial type (repeat/switch). We report RT averages based on raw values, but the reported analyses and model results are based on log RTs. The final lme model (see Appendix B) includes random intercepts of participant with random slopes of task and trial type (and their interaction) and the random intercepts of item with the random slope of task.

Results show a significant main effect of task type ($\beta = -0.10$, $SE = 0.01$, $t = -8.51$, $p < .001$), with naming in the voluntary task ($M = 845$, $SD = 213$) being faster than naming in the cued task ($M = 1028$, $SD = 309$). There was also a main effect of trial type ($\beta = -0.04$, $SE = 0.01$, $t = -11.26$, $p < .001$), with faster naming on repeat trials ($M = 912$, $SD = 270$) than switch trials ($M = 971$, $SD = 293$), reflecting a general switch cost. We also observed an interaction between task and trial type ($\beta = 0.02$, $SE = 0.01$, $t = 6.43$, $p < .001$).

To further investigate this interaction, reflecting a difference in switch cost in the cued versus voluntary task, we analysed the tasks separately. Both the voluntary task ($\beta = -0.02$, $SE = 0.004$, $t = -4.35$, $p < .001$) and the cued task ($\beta = -0.05$, $SE = 0.004$, $t = -12.35$, $p < .001$) showed a switch cost. However, this cost was smaller in the voluntary task ($M_{\text{switch}} = 865$, $SD_{\text{switch}} = 225$;

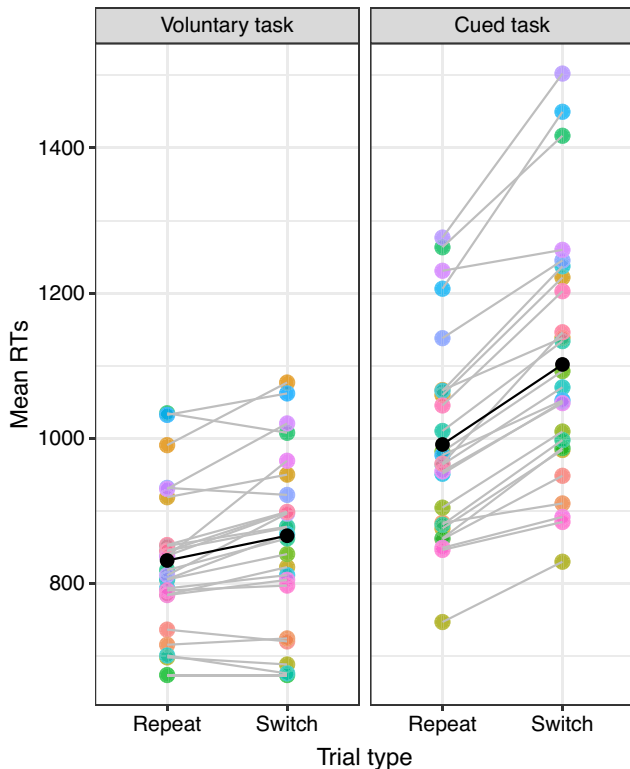


Figure 2. Mean naming latencies (per participant) as a function of task type and trial type. Each dot represents one participant. Black dots and lines represent overall mean per condition. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/YWAVN>.

$M_{\text{repeat}} = 832$, $SD_{\text{repeat}} = 205$) than in the cued task ($M_{\text{switch}} = 1102$, $SD_{\text{switch}} = 313$; $M_{\text{repeat}} = 992$, $SD_{\text{repeat}} = 301$). This means that while the voluntary switch cost was significantly smaller than the cued switch cost, voluntary switching was still costly.

3.1.2 ERP analysis

Figure 3 shows the average ERPs and topographies for the cued and voluntary switch effect, as well as the general switch effect collapsed over tasks. As expected, there was a numerical difference between the conditions, with a slightly larger N2 peak in switch trials than repeat trials. However, the cluster-based permutation test of the general switch effect showed this difference between repeat and switch trials overall was not significant (no clusters were detected by the clustering procedure) in the N2 time window (i.e., 180 to 300 ms). Similarly, there was no significant N2 effect for switch compared to repeat trials in the cued task ($p = .581$) or in the voluntary task ($p = .422$) in task-specific tests. The permutation test for the interaction between voluntary and cued switch cost was not significant either ($p = .759$).

Figure 4 (left panel) shows the average ERPs of the cued and voluntary tasks in a representative frontal and posterior cluster. While there were no N2 switch effects or differences between the switch costs for cued and voluntary trials, the cluster-based permutation test for the task effect across switch and non-switch trials did reveal a significantly larger N2 for trials in the cued task than voluntary trials ($p = .012$). The effect was present in the entire time window of 180–300 ms post-stimulus onset and was mostly noted at posterior sites. We performed a Laplacian transform on the ERP data (using the spline method, Perrin et al., 1989) to increase the focality of the scalp topography of this effect, clarifying its

generation by more posterior sources. Figure 4 (right panel) shows the topographical map of the most prominent cluster, confirming its predominantly posterior bilateral distribution. We will address this finding in the discussion.

3.1.3 Time-resolved power analysis

Figure 5 presents the average relative power changes for the switch effect (relative power changes for switch versus repeat trials) for the voluntary and the cued task between 0 and 700 ms post-stimulus onset. Relevant effects here are the differences in theta-band power between switch and repeat trials, which is indicated in the very bottom parts (the 4–8 Hz band) of the top panels in the figure. A cluster-based permutation test of the general switch effect (collapsed over task type) revealed no statistically significant difference in theta-band power between repeat and switch trials ($p = .464$). As can be seen in the top left panel of figure 5, there is a theta power increase in switch compared to repeat trials in the cued task, but this difference is not statistically significant ($p = .542$). This non-significant difference is not visible for the voluntary task ($p = .534$). However, the permutation test showed no significant between-task interaction when inspecting the difference between the cued and voluntary switch effects (no clusters were detected). Finally, there was also no significant difference between the theta-band power in midfrontal electrodes for the cued task in comparison to the voluntary task: no clusters were detected by the clustering procedure.

To examine whether there was any difference in other frequency bands than theta, a non-targeted cluster-based permutation scoured the time-frequency space of the 15 midfrontal channels of interest for all frequencies up to 30 Hz. No significant effects were found.

3.1.4 Summary: planned analyses

Speakers responded faster on repeat than on switch trials in both voluntary and cued switching tasks. This switch cost was significantly smaller when participants were free to decide when to switch in comparison to switching on instruction. However, there were no significant N2 or midfrontal theta switch effects in any condition. The task effect in the N2 window was significant: we observed a larger N2 on cued compared to voluntary trials, with the effect being most pronounced at posterior sites.

3.2 Exploratory analyses

To determine whether response latencies for the various conditions differed between Dutch and English, language of trial was added to the lme model to investigate a potential behavioural language effect (see Appendix B). Table 2 shows average RTs (and standard deviations) for each condition for this analysis. In addition to all previously described effects, there was a significant main effect of language, with naming in English being faster than naming in Dutch ($\beta = 0.03$, $SE = 0.003$, $t = 12.12$, $p < .001$). There was also an interaction between task type and language ($\beta = -0.01$, $SE = 0.002$, $t = -6.67$, $p < .001$), with the language effect being larger in the cued task than the voluntary task. The interaction between trial type and language was also significant ($\beta = 0.01$, $SE = 0.002$, $t = 2.70$, $p = .007$). The switch cost (the difference between switch and repeat trials) was bigger in English than in Dutch trials, although it was still significant in both languages separately. Finally, the three-way interaction between task type, trial type and language was also significant ($\beta = -0.01$, $SE = 0.002$, $t = -2.92$, $p = .004$). The difference between the switch cost in Dutch

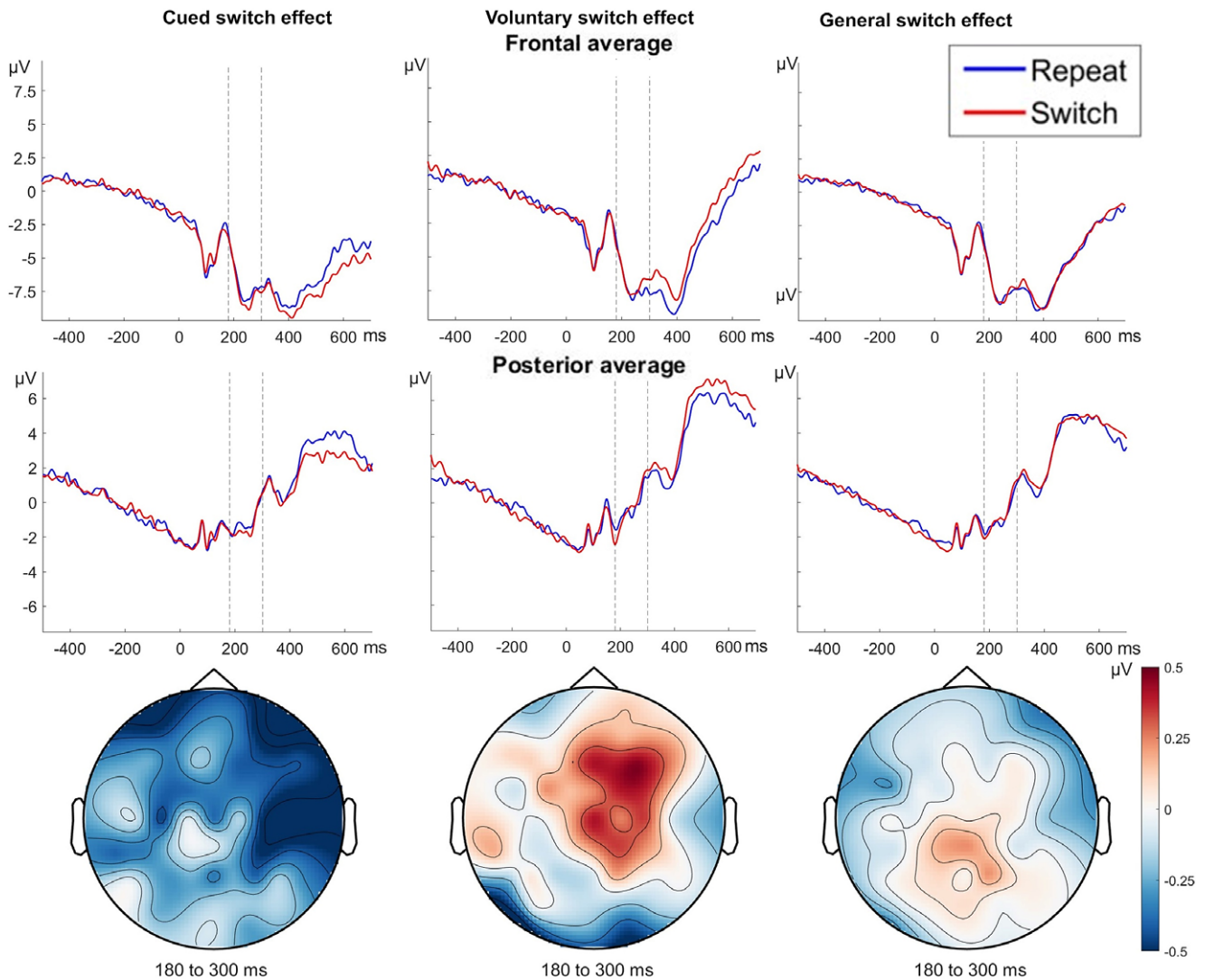


Figure 3. Stimulus-locked ERPs and topographies for switch versus repeat trials, based on a frontal cluster of eight electrodes (FC1, FCz, FC2, Fz, F3, F2, FC3 and F1) and a posterior cluster of eight electrodes (CP2, Pz, P3, CP1, P2, CPz, P1 and CP3). Right panels represent the general switch effect while the left and middle panels show the switch effect in the cued and voluntary tasks respectively. The dashed lines mark the time window of interest (180 to 300 ms). Topographies show the difference between repeat and switch trials (computed as switch – repeat). Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/YWAVN>.

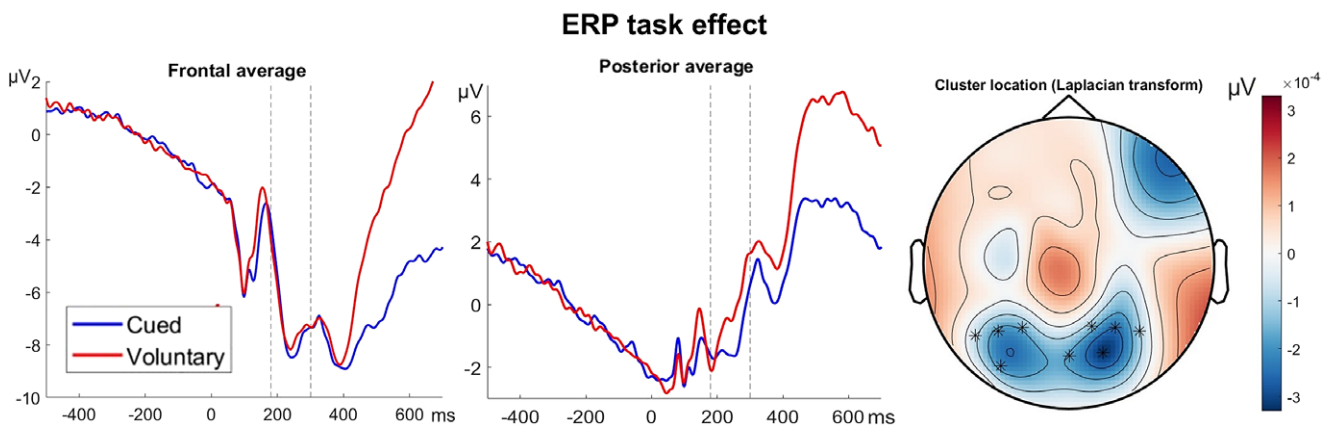


Figure 4. Stimulus-locked ERPs showing the task effect and topographical map showing the location of the cluster associated with the significant effect. The left panel shows a representative frontal channel cluster of eight electrodes (FC1, FCz, FC2, Fz, F3, F2, FC3 and F1), and the middle panel shows a representative posterior cluster of eight electrodes (CP2, Pz, P3, CP1, P2, CPz, P1 and CP3) for the task effect. Dashed lines indicate the time window of interest (180 to 300 ms). Right: topographical map showing the location of the observed effect after the Laplacian transform. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/YWAVN>.

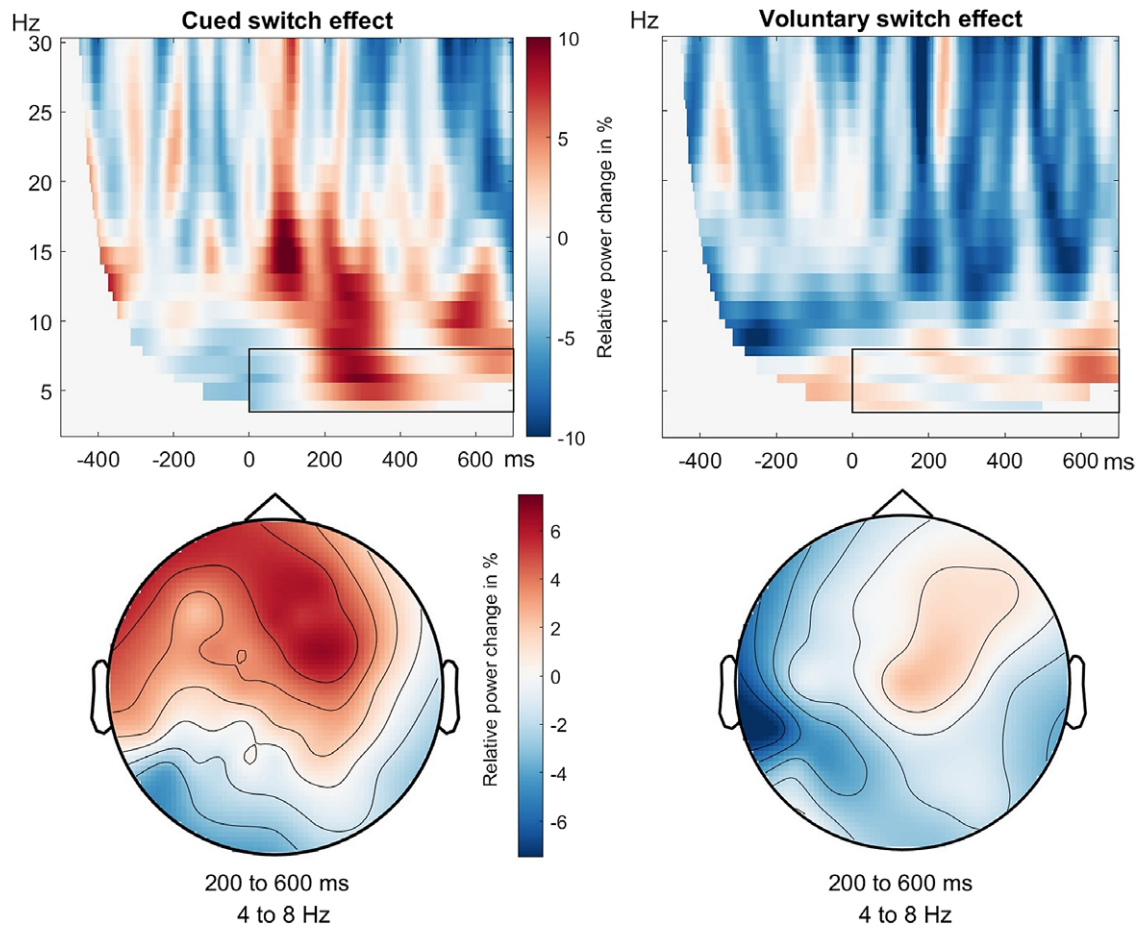


Figure 5. Time-resolved power of the switch effect (power in switch trials – repeat trials, normalised by their average) and topographical maps of switch effects. Top: the left panel represents the switch effect in the cued task, and the right panel represents the effect in the voluntary task. These graphs show the average over the following channels: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2 and C4. The areas marked in black show the time and frequency window of interest for the statistical analysis (0–700 ms, 4–8 Hz). Bottom: topographical maps showing the switch effect in the theta band (4–8 Hz) between 200 and 600 ms after stimulus presentation. Reproduced with permission from the authors from <https://doi.org/10.17605/OSF.IO/YWAVN>.

and English was only present in the cued task ($M_{\text{switch cost English}} = 133$, SD of difference = 404; $M_{\text{switch cost Dutch}} = 87$, SD of difference = 456). In the voluntary task, the difference was not significant, although it is interesting to note that it was numerically reversed, with a larger switch cost in Dutch than English ($M_{\text{switch cost English}} = 30$, SD of difference = 299; $M_{\text{switch cost Dutch}} = 37$, SD of difference = 310). Overall, these exploratory results show that participants were faster to respond on English than Dutch trials, a difference which was larger in the cued than the voluntary task.

Post-hoc, it also became clear that task order could have been a factor to influence switch costs (see discussion). Task order was added to the non-language-specific lme model (see section 3.1.1). There was no effect of task order on RTs ($\beta = -.03$, $SE = 0.02$, $t = 1.13$, $p = .269$), nor was there an interaction between task order and trial type ($\beta = -0.001$, $SE = 0.003$, $t = -0.12$, $p = .909$), meaning task order did not influence switch cost.

4. Discussion

The current study investigated the behavioural and electrophysiological effects of language control while switching in a cued and voluntary context. Participants were significantly faster at naming pictures and switching languages in voluntary versus cued contexts,

but we did not obtain evidence for switch effects in either task in the ERPs or midfrontal theta. There was, however, an ERP effect of tasks, with a more extreme posterior negative peak in the cued task than the voluntary task.

4.1 Behavioural effects in voluntary and cued switching

The finding of a smaller voluntary than cued switch cost behaviourally is compatible with findings from a number of studies (Experiment 2, Gollan et al., 2014; Jevtović et al., 2020) but contradicts some others, in which behavioural voluntary and cued switch costs were similar (de Bruin et al., 2018; Gollan et al., 2014). In a cued language-switching context, language control processes such as response inhibition, goal maintenance, conflict monitoring, and cue detection are argued to be required for using and switching between languages based on external cues (e.g., Green & Abutalebi, 2013). When speakers are conversing with an interlocutor who shares their multiple languages, there could be less need for these processes as they can choose whichever language is most accessible at any given time. Our results follow this logic: it seems to be behaviourally easier and faster to use two languages and switch when the speaker chooses to, compared to switching based on external cues.

Table 2. Reaction times (ms) in the voluntary and cued tasks, as used for the exploratory analysis that included language in the model. Showing means and standard deviations by task, trial type and language. Note: the overall summary means for language are based on an uneven distribution of trial numbers between trial types

Task type	Trial type	Language	Mean (SD)
Repeat		English	904 (259)
		Dutch	964 (297)
	Switch	English	951 (282)
		Dutch	991 (301)
	Repeat	English	880 (242)
		Dutch	949 (293)
Cued		English	981 (286)
		Dutch	1076 (324)
	Switch	English	1070 (303)
		Dutch	1134 (320)
	Repeat	English	937 (267)
		Dutch	1047 (322)
Voluntary		English	836 (209)
		Dutch	855 (218)
	Switch	English	855 (222)
		Dutch	877 (228)
	Repeat	English	825 (201)
		Dutch	839 (209)

We furthermore show that these findings extend to Dutch-English bilinguals with a high proficiency in both languages. Participants in the current study learned English at an average age of 10 and though they use a considerable amount of English in their daily life, all report speaking Dutch at home and in (most) social situations. This indicates that the ease of voluntary switching between languages is not restricted to those who are used to dense language-switching contexts in everyday situations.

While voluntary switching was more efficient than cued switching, we still observed a voluntary switch cost, as participants were slower to respond to switch than repeat trials. Studies reporting the absence of a switch cost (Kleinman & Gollan, 2016; Zhu *et al.*, 2022) often induce language switching based on lexical accessibility by creating consistency for each item in the language used to name it (either naturally, as in Zhu *et al.*, or artificially, as in Kleinman & Gollan). When such instructions or designs are not used, lexical access is still one of the key variables driving language choice and switching (e.g., de Bruin *et al.*, 2018). However, other top-down processes might also play a role. For instance, bilinguals might use a default language (e.g., Grosjean, 1998) that they prefer to switch back to even if this is not the most efficient choice. Their language choice and switching might also depend on previously named items, the context or interlocutor (including the switching behaviour of the experiment leader) and switching strategies to make sure both languages are used in the task.

Bilinguals naturally differ in their approach towards language use in voluntary tasks (e.g., de Bruin *et al.*, 2018; Gollan *et al.*, 2014). Previous studies examining whether switch costs are related to the speaker's approach, done within bilinguals and in the absence of further task instructions, have shown mixed results. While some

suggest bilinguals who name each word consistently in the same language show smaller switching costs (e.g., Gollan *et al.*, 2014), others have found naming consistency to be associated with overall naming times but not switch costs (e.g., de Bruin *et al.*, 2018). This might be because all bilinguals switch for other reasons than pure lexical access in some cases, including in daily life, with corpus studies also showing switching costs (Fricke *et al.*, 2016). Thus, even when language choice is free, the influence of external factors and a certain degree of top-down control might remain present during switching.

Besides the switch effects, our exploratory analyses indicate that speakers responded more quickly on trials in English (L2) than in Dutch (L1). This reverse dominance effect has been observed in many studies (Christoffels *et al.*, 2007; Verhoef *et al.*, 2010; Zheng *et al.*, 2020). It might be caused by global inhibition of the dominant language during the entire task to facilitate production in the non-dominant language. If bilinguals apply this inhibition too strongly, it can result in relatively slower L1 than L2 responses. This pattern, while present in both tasks, was strongest in the cued task. This could be another argument for the hypothesis that more top-down control is used in the cued task, even if it is present in both: global inhibition of the L1 compared to the L2 may be less strong in the voluntary than the cued task as less inhibition is needed generally. However, in the absence of a baseline single-language condition, it is not possible to conclude whether these effects reflect language control specifically or rather more general differences in lexical retrieval speed.

4.2 Electrophysiological signatures of language switching

We examined two neural markers of cognitive control to compare switch effects in voluntary and cued language switching: the N2 ERP component and midfrontal theta power. A number of previous studies have found a more extreme N2 peak for cued switch versus repeat trials (Declerck *et al.*, 2021; Jackson *et al.*, 2001; Kang *et al.*, 2020; Verhoef *et al.*, 2010; Zheng *et al.*, 2020). We also expected to find a midfrontal theta switch effect based on more recent work (Cui *et al.*, 2024), in which a midfrontal theta cued switch effect was consistently observed in several datasets. We failed to replicate both effects in the present study: switching between languages did not significantly affect the N2 peak when compared to trials in which the language remained the same. Similarly, we found a midfrontal theta modulation in the expected direction, but this was not significant. This was the case for switching between languages in both voluntary and cued contexts.

Before discussing the absence of an electrophysiological switch effect, it is worth considering that the ERP data showed a main effect of task type, with a more extreme N2 peak in cued than voluntary trials. This suggests that the two tasks were, in some ways, processed differently by the participants. However, the posterior location of this effect suggests these differences do not reflect language control, considering N2 effects reflecting cognitive control in non-linguistic task switching are almost always reported in frontocentral regions (Gajewski *et al.*, 2010; Swainson *et al.*, 2003). We believe this more extreme peak in the N2 time window for the cued compared to the voluntary task might not be an N2 effect related to top-down control but rather a visual (attentional) effect caused by the different levels of depth of processing required for the colour cues in the tasks. In the cued task, participants needed to process the visual cue to decide which language to use. Visual processing of the background colours in the voluntary task could be

much shallower or even absent because the background colours of the images did not reveal any task-relevant information.

It is noteworthy that the task effect in the present study has a similar latency and location to the visual awareness negativity (VAN) component. This has previously been reported when comparing trials that require conscious visual processing to unconscious visual processing trials (Busch et al., 2010; Derda et al., 2019; Ojanen et al., 2003). Notably, the VAN's latency overlaps with the time window for an effect of language control (the N2), which many would argue is a later cognitive process. The larger discussion of how time windows of ERP components reflect true timings of cognitive processes is beyond the scope of this paper. Nevertheless, the main task effect suggests our study was sensitive to differences between tasks, but this ERP modulation is unlikely to reflect differences in top-down language control, in line with the absence of (differences between) switch costs, to which we turn next. This difference in cue processing depth between tasks can be addressed in future studies by using paradigms that do not require cues in either the voluntary or forced switching tasks, or paradigms that can separate cue and stimulus processing (e.g., Verhoef et al., 2010).

One potential explanation for the absence of N2 switching costs in both tasks is the rather inconsistent nature of the N2 as a language control index. Some of its aspects are still debated in the literature, while other studies investigating the N2 as an electrophysiological switch cost find it absent (e.g., Christoffels et al., 2007; Zhu et al., 2022). There are two common functional interpretations of the N2 switch effect: it either indicates overcoming inhibition of the target language or active inhibition of/disengagement from the non-target language. These different interpretations are also related to the possible scalp distributions of the effect: some studies have found a frontal N2 switch effect (Declerck et al., 2021; Jackson et al., 2001; Kang et al., 2020), while others found a posterior N2 switch effect (Verhoef et al., 2010; Zheng et al., 2020). It is also quite common for the effect to only be found in switches to the L2 in language-specific analyses (Jackson et al., 2001; Verhoef et al., 2010; Zheng et al., 2020).¹ Overall, the findings of posterior as well as anterior N2 switch effects go to show that consensus on the location and interpretation of the effect has not yet been reached. It remains an open question whether these inconsistencies in the literature are a sign of a noisy effect that is nevertheless meaningful when it can be appropriately measured, or an indication that the N2 component is not a marker of cognitive control in language switching the way scholars have hypothesised.

The absence of EEG switch effects in the current study may also have been caused by population factors leading to a smaller effect size. One of these could be language proficiency in the L2: most studies that find a significant N2 switch effect used participants whose L2 proficiency was considerably lower (about intermediate level) than ours. We explored the potential influence of proficiency by comparing our participants' LexTALE scores to Zheng et al.'s (2020), whose data show an N2 and midfrontal theta switch effect. The participants in the current study ($M = 87.0$, $SD = 10.6$) scored significantly higher than Zheng et al.'s participants ($M = 78.2$, $SD = 12.2$, $t(50) = 3.00$, $p = .004$) on the LexTALE, indicating higher L2 proficiency. The current participants frequently used their L2 in their daily life as part of their degree or work, explaining their higher English proficiency. Our participants also reported switching between Dutch and English in everyday situations (across scores for within a day, within a conversation, and within

a sentence) more frequently ($M = 5.65$, $SD = 1.67$) than Zheng et al.'s participants ($M = 4.33$, $SD = 2.03$, $t(53) = 2.63$, $p = .011$) on a scale of 1–10. These variables could possibly have affected participants' language control during an experimental task. Indeed, both other studies that found no N2 switch effect (Christoffels et al., 2007; Zhu et al., 2022) used participants that were highly proficient in both languages or switched languages frequently in their daily life. This difference in proficiency and code-switching habits might have decreased the effect size of the electrophysiological switch effect in the cued task. It is possible that more balanced activation of both languages requires less control during language switching and leads to smaller switching costs – too small to detect in the EEG data with our sample size.

The second difference in design to most cued language-switching tasks is the addition of the voluntary task within the same session. In theory, this combination of tasks could potentially have decreased the effect size of the cued electrophysiological switch effects. Although one could analyse the participants in the current study who started with the cued task first to determine whether the nature and order of the task combinations affected the switch cost, this analysis would only be based on 13 participants. Decreasing the sample size affects power, preventing conclusive claims about this issue. Of note, in the RTs, task order did not significantly affect the switch costs.

The proficiency and design related differences above may also partially explain the absence of a cued switch effect in midfrontal theta oscillations. Midfrontal theta switch costs have previously been observed in non-linguistic task switching (Cooper et al., 2017) as well as cued language switching (Cui et al., 2024). Still, oscillation-based methodologies are rather novel in the study of language control, which consequently leaves the effect sizes yet unknown. Given the effect of participants' L2 proficiency outlined above, this issue will require further consideration in future studies.

Finally, there was a lack of electrophysiological switch effects in the voluntary task. While the previously discussed factors also apply here, this absence could theoretically also be due to the difference in nature between the tasks. Following the logic of the ACH, switching between languages as desired could be less demanding for various cognitive control processes. Our behavioural results do support this theory, showing that there is a smaller switch cost in the voluntary than the cued task. Electrophysiological effect sizes in the voluntary task are likely to be smaller than the already undetectable effects in the cued task, which could explain their absence. Alternatively, we might not have detected an electrophysiological switch cost in voluntary switching because there is none: if language switching is mostly based on bottom-up lexical access (rather than top-down control), there might be few differences in the specific electrophysiological processes of language control measured in this study. This is likely to be subject to individual differences in the same way the behavioural voluntary switch cost could be, based on the speaker's proficiency and habits, the context, or possible switching strategies.

Of note, there are other cases where electrophysiological and behavioural measures do not converge: some language production studies find behavioural effects where there are none in the electrophysiology (e.g., Piai, Roelofs, & Maris, 2014), while others find EEG but no behavioural effects (e.g., Jiao et al., 2022).

5. Conclusion

While cued language switching has been regularly investigated, relatively few studies have looked at switching between languages in a voluntary context, even though voluntary switching

¹Exploratory analyses of the current study's cued switch effect in English showed no N2 effects, but were likely underpowered.

is a common phenomenon in natural multilingual situations. The present study explored voluntary and cued language switching in highly proficient Dutch-English late bilinguals and examined how the behavioural and neural language control processes varied across tasks. Behaviourally, we found that switching between languages is not as effortful as in cued language-switching studies. However, the missing evidence for electrophysiological modulation of switch costs suggests behavioural switch costs cannot always be easily mapped onto neural switch costs. The effect size of neural switch effects may be altered by design and population changes while behavioural effects remain more robust.

Data availability. Anonymised data, study materials and code are available from the Radboud Data Repository at <https://doi.org/10.34973/w0z9-p892>.

Acknowledgements. This study was funded by the Gravitation Grant (024.001.006) of the Language in Interaction Consortium from the Netherlands Organization for Scientific Research (NWO).

Competing interests. The author(s) declare none.

References

- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/JSS.V067.I01>
- Blanco-Elorrieta, E., & Pyllkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *The Journal of Neuroscience*, *37*(37), 9022–9036. <https://doi.org/10.1523/JNEUROSCI.0553-17.2017>
- Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer (Version 6.1.41) [Computer software]. Praat.org. <http://www.praat.org/>
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. <https://doi.org/10.1016/j.tics.2011.12.010>
- Busch, N. A., Fründ, I., & Herrmann, C. S. (2010). Electrophysiological evidence for different types of change detection and change blindness. *Journal of Cognitive Neuroscience*, *22*(8), 1852–1869. <https://doi.org/10.1162/JOCN.2009.21294>
- Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. *Cerebral Cortex*, *22*(11), 2575–2586. <https://doi.org/10.1093/cercor/bhr332>
- Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. <https://doi.org/10.1016/J.TICS.2014.04.012>
- Champely, S. (2020). *pwr: Basic functions for power analysis* (R package version 1.3–0) [Computer software]. <https://CRAN.R-project.org/package=pwr>
- Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*(1), 192–208. <https://doi.org/10.1016/J.BRAINRES.2007.01.137>
- Cohen, M. X., & Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *Journal of Neurophysiology*, *110*(12), 2752–2763. <https://doi.org/10.1152/JN.00479.2013/ASSET/IMAGES/LARGE/Z9K0241322250006.JPG>
- Cooper, P. S., Karayanidis, F., McKewen, M., McLellan-Hall, S., Wong, A. S. W., Skippen, P., & Cavanagh, J. F. (2019). Frontal theta predicts specific cognitive control-induced behavioural changes beyond general reaction time slowing. *NeuroImage*, *189*, 130–140. <https://doi.org/10.1016/j.neuroimage.2019.01.022>
- Cooper, P. S., Wong, A. S. W., McKewen, M., Michie, P. T., & Karayanidis, F. (2017). Frontoparietal theta oscillations during proactive control are associated with goal-updating and reduced behavioral variability. *Biological Psychology*, *129*, 253–264. <https://doi.org/10.1016/J.BIOPSYCHO.2017.09.008>
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511. <https://doi.org/10.1016/J.JML.2004.02.002>
- Cui, N., Piai, V., & Zheng, X. Y. (2024). Domain-general cognitive control processes in bilingual switching: Evidence from midfrontal theta oscillations. *European Journal of Neuroscience*, *60*(5), 4813–4829. <https://doi.org/10.1111/ejn.16466>
- de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28–43. <https://doi.org/10.1016/J.JML.2018.07.005>
- de Bruin, A., & Xu, T. (2022). Language switching in different contexts and modalities: Response-stimulus interval influences cued-naming but not voluntary-naming or comprehension language-switching costs. *Bilingualism: Language and Cognition*, 1–14. <https://doi.org/10.1017/S1366728922000554>
- Declerck, M., Meade, G., Midgley, K. J., Holcomb, P. J., Roelofs, A., & Emmorey, K. (2021). On the connection between language control and executive control—An ERP study. *Neurobiology of Language*, 1–19. https://doi.org/10.1162/NOL_A_00032
- Derda, M., Koculak, M., Windey, B., Gociewicz, K., Wierchoń, M., Cleermans, A., & Binder, M. (2019). The role of levels of processing in disentangling the ERP signatures of conscious visual processing. *Consciousness and Cognition*, *73*, 102767. <https://doi.org/10.1016/J.CONCOG.2019.102767>
- Donkers, F. C. L., & Van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition*, *56*(2), 165–176. <https://doi.org/10.1016/J.BANDC.2004.04.005>
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. <https://doi.org/10.1080/17470218.2017.1310261>
- Eisma, J., Rawls, E., Long, S., Mach, R., & Lamm, C. (2021). Frontal midline theta differentiates separate cognitive control strategies while still generalizing the need for cognitive control. *Scientific Reports*, *11*(1), 14641. <https://doi.org/10.1038/s41598-021-94162-z>
- Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, *91*, 181–201. <https://doi.org/10.1016/J.JML.2016.04.003>
- Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, *89*, 110–137. <https://doi.org/10.1016/J.JML.2015.10.001>
- Gajewski, P. D., Kleinsorge, T., & Falkenstein, M. (2010). Electrophysiological correlates of residual switch costs. *Cortex*, *46*(9), 1138–1148. <https://doi.org/10.1016/J.CORTEX.2009.07.014>
- Gollan, T. H., Kleinman, D., & Wierenga, C. E. (2014). What’s easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General*, *143*(6), 2167. <https://doi.org/10.1037/A0038006>
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. <https://doi.org/10.1017/S1366728998000133>
- Green, D. W., & Abutalebi, J. (2013). The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. <https://doi.org/10.1080/20445911.2013.796377>
- Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149. <https://doi.org/10.1017/S136672899800025X>
- Heikoop, K. W., Declerck, M., Los, S. A., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language switching. *Bilingualism: Language and Cognition*, *19*(5), 921–927. <https://doi.org/10.1017/S1366728916000456>
- Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition*, *4*(2), 169–178. <https://doi.org/10.1017/S1366728901000268>
- Jevtović, M., Duñabeitia, J. A., & De Bruin, A. (2020). How do bilinguals switch between languages in different interactional contexts? A comparison between

- voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, **23**(2), 401–413. <https://doi.org/10.1017/S1366728919000191>
- Jiao, L., Gao, Y., Schwieter, J. W., Li, L., Zhu, M., & Liu, C. (2022). Control mechanisms in voluntary versus mandatory language switching: Evidence from ERPs. *International Journal of Psychophysiology*, **178**, 43–50. <https://doi.org/10.1016/j.ijpsycho.2022.06.005>
- Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and Clinical Neurophysiology*, **82**(6), 477–482. [https://doi.org/10.1016/0013-4694\(92\)90054-L](https://doi.org/10.1016/0013-4694(92)90054-L)
- Kang, C., Ma, F., Li, S., Kroll, J. F., & Guo, T. (2020). Domain-general inhibition predicts the intensity of inhibition on non-target language in bilingual word production: An ERP study. *Bilingualism: Language and Cognition*, **23**(5), 1056–1069. <https://doi.org/10.1017/S1366728920000085>
- Keuleers, E., & Brysbaert, M. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, **42**(3), 643–650. <https://doi.org/10.3758/BRM.42.3.643>
- Kleinman, D., & Gollan, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science*, **27**(5), 700–714. <https://doi.org/10.1177/0956797616634633>
- Krott, A., Medaglia, M. T., & Porcaro, C. (2019). Early and late effects of semantic distractors on electroencephalographic responses during overt picture naming. *Frontiers in Psychology*, **10**(MAR), 696. <https://doi.org/10.3389/FPSYG.2019.00696/BIBTEX>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, **44**(2), 325–343. <https://doi.org/10.3758/S13428-011-0146-0>
- Li, S., Botezatu, M. R., Zhang, M., & Guo, T. (2021). Different inhibitory control components predict different levels of language control in bilinguals. *Memory & Cognition*, **49**(4), 758–770. <https://doi.org/10.3758/s13421-020-01131-4>
- Lüdtke, D. (2021). *sjPlot: Data visualization for statistics in social science* (2.8.10) [Computer software]. <https://CRAN.R-project.org/package=sjPlot>
- 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. <https://doi.org/10.1016/J.CLINPH.2004.03.031>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, **164**(1), 177–190. <https://doi.org/10.1016/J.JNEUMETH.2007.03.024>
- Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning Memory and Cognition*, **29**(3), 362–372. <https://doi.org/10.1037/0278-7393.29.3.362>
- Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, **40**(1), 25–40. <https://doi.org/10.1006/JMLA.1998.2602>
- Ojanen, V., Revonsuo, A., & Sams, M. (2003). Visual awareness of low-contrast stimuli is reflected in event-related brain potentials. *Psychophysiology*, **40**(2), 192–197. <https://doi.org/10.1111/1469-8986.00021>
- 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**. <https://doi.org/10.1155/2011/156869>
- Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, **72**(2), 184–187. [https://doi.org/10.1016/0013-4694\(89\)90180-6](https://doi.org/10.1016/0013-4694(89)90180-6)
- Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology*, **19**(3), 395–416. <https://doi.org/10.1080/09541440600758812>
- 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**(2), e88674. <https://doi.org/10.1371/JOURNAL.PONE.0088674>
- Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*, **53**, 146–156. <https://doi.org/10.1016/j.neuropsychologia.2013.11.014>
- Piai, V., & Zheng, X. (2019). Speaking waves: Neuronal oscillations in language production. *Psychology of Learning and Motivation—Advances in Research and Theory*, **71**, 265–302. <https://doi.org/10.1016/BS.PLM.2019.07.002>
- R Core Team (2021). *R: A Language and Environment for Statistical Computing* (Version 4.1.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, **37**(4), 473–484. <https://doi.org/10.1111/1469-8986.3740473>
- Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M. (2017). Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex*, **92**, 289–303. <https://doi.org/10.1016/J.CORTEX.2017.04.017>
- Singman, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2023). *afex: Analysis of Factorial Experiments* (R package version 1.3–0) [Computer software]. <https://CRAN.R-project.org/package=afex>
- Swanson, R., Cunnington, R., Jackson, G. M., Rorden, C., Peters, A. M., Morris, P. G., & Jackson, S. R. (2003). Cognitive control mechanisms revealed by ERP and fMRI: Evidence from repeated task-switching. *Journal of Cognitive Neuroscience*, **15**(6), 785–799. <https://doi.org/10.1162/089892903322370717>
- Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, **52**(8), 997–1009. <https://doi.org/10.1111/psyp.12437>
- Trevartha, K. M., & Phillips, N. A. (2013). Detecting self-produced speech errors before and after articulation: An ERP investigation. *Frontiers in Human Neuroscience*, **7**, 763. <https://doi.org/10.3389/FNHUM.2013.00763/BIBTEX>
- Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, **38**(4), 584–589. <https://doi.org/10.3758/BF03193889>
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, **67**(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, **110**(1), 84–99. <https://doi.org/10.1016/J.COGNITION.2008.10.013>
- Verhoef, K., Roelofs, A., & Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience*, **22**(8), 1832–1843. <https://doi.org/10.1162/JOCN.2009.21291>
- Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data—a practical approach. *Journal of Neuroscience Methods*, **250**, 34–46. <https://doi.org/10.1016/j.jneumeth.2014.08.002>
- Zheng, X., Roelofs, A., Erkan, H., & Lemhöfer, K. (2020). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study. *Neuropsychologia*, **140**, 107387. <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2020.107387>
- Zhu, J. D., Blanco-Elorrieta, E., Sun, Y., Szakay, A., & Sowman, P. F. (2022). Natural vs forced language switching: Free selection and consistent language use eliminate significant performance costs and cognitive demands in the brain. *NeuroImage*, **247**, 118797. <https://doi.org/10.1016/J.NEUROIMAGE.2021.118797>

Appendix A: stimuli

Item number	English name	Dutch name	Set	Item number	English name	Dutch name	Set
1	umbrella	paraplu	A	21	chicken	kip	B
2	mountain	berg	A	22	fridge	koelkast	B
3	painting	schilderij	A	23	mirror	spiegel	B
4	desert	woestijn	A	24	newspaper	krant	B
5	onion	ui	A	25	butterfly	vlinder	B
6	pencil	potlood	A	26	squirrel	eekhoorn	B
7	lettuce	sla	A	27	mushroom	paddenstoel	B
8	candle	kaars	A	28	arrow	pijl	B
9	hippo	nijlpaard	A	29	waiter	ober	B
10	wall	muur	A	30	car	auto	B
11	eye	oog	A	31	dog	hond	B
12	meat	vlees	A	32	doll	pop	B
13	plate	bord	A	33	key	sleutel	B
14	duck	eend	A	34	horse	paard	B
15	knife	mes	A	35	leg	been	B
16	spoon	lepel	A	36	dress	jurk	B
17	branch	tak	A	37	belt	riem	B
18	frog	kikker	A	38	ant	mier	B
19	flower	bloem	A	39	corn	maïs	B
20	tree	boom	A	40	axe	bijl	B

Appendix B: linear mixed-effects models

Variables: task is a task type (voluntary/cued), SwiSta_f is a trial type (switch status: switch/repeat), Subject_nr is the participant number and item is the stimulus image. Language_f is the language of the trial. Data_correct contains all trials used for the analysis, after pre-processing (see Section 2.7.1 for exclusion criteria).

Main model

#Final model as reported, with log-corrected RTs.

```
Final_model = mixed (log_Corrected_RT ~ task_f * SwiSta_f +
  (1 + task_f * SwiSta_f | Subject_nr) + (1 + task_f |
  Item_no),
  data = data_correct,
  control = lmerControl(optimizer = "bobyqa",
    optCtrl = list(maxfun = 1000000)))
```

Cued task model

```
model_cued = mixed (log_Corrected_RT ~ SwiSta_f +
  (1 + SwiSta_f | Subject_nr) + (1 | Item_no),
```

```
data = cued_switch,
control = lmerControl(optimizer = "bobyqa",
  optCtrl = list(maxfun = 1000000)))
```

Voluntary task model

```
model_vol = mixed (log_Corrected_RT ~ SwiSta_f +
  (1 + SwiSta_f | Subject_nr) + (1 | Item_no),
  data = vol_switch,
  control = lmerControl(optimizer = "bobyqa",
    optCtrl = list(maxfun = 1000000)))
```

Model including language effect

```
model_language = mixed (log_Corrected_RT ~ task_f * SwiSta_f *
  language_f +
  (1 + task_f * SwiSta_f | Subject_nr) + (1 + task_f |
  Item_no),
  data = data_correct,
  control = lmerControl(optimizer = "bobyqa",
    optCtrl = list(maxfun = 1000000)))
```