Voluntary language switching: When and why do bilinguals switch between their languages?

Angela de Bruin[1], Arthur G. Samuel[1,2,3], & Jon Andoni Duñabeitia[4,1]

[1]Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain

[2] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

[3] Department of Psychology, Stony Brook University, Stony Brook, NY, USA

[4] Facultad de Lenguas y Educación, Universidad Nebrija, Madrid, Spain

Corresponding author:

Angela de Bruin

Basque Center on Cognition, Brain and Language

Paseo Mikeletegi 69

20009 Donostia, Spain

Phone: +34 943 309 300 Ext. (225)

Email: a.debruin@bcbl.eu

Abstract

Bilingual language switching has been studied extensively in cued picture naming paradigms, instructing bilinguals when to switch between languages. However, in daily life, bilinguals often switch freely, without external instruction. This study examined when and why bilinguals switch voluntarily. Spanish-Basque bilinguals frequently switched between their languages and their language choice was related to the ease of lexical access. Words that were slow to be accessed in Basque were more often named in Spanish and vice versa. In terms of response times, switching costs were observed not only in the cued but also in the voluntary task. However, while cued switching showed a mixing cost (reflecting the cost associated with using two languages rather than one), a mixing benefit was observed for the voluntary task. This suggests that voluntarily using two languages may be less costly than having to stay in one language.

Keywords: Bilingualism; voluntary language switching; cued language switching; lexical access; language control

Word count (excl. references and appendices): 12.900

Introduction

Bilinguals frequently switch between their languages in daily life. Sometimes, language switching is imposed by the circumstances and interlocutors. For instance, when different languages are used with different interlocutors, monitoring of the context is needed to select the appropriate target language and to switch at the right time. Laboratory studies have focused on this type of language switching by using cued picture or digit naming tasks in which cues (e.g., the flag of a country) indicate the language that needs to be used. In daily life, however, language switching does not always follow cues but can also take place freely. When both conversational partners speak the same languages, bilinguals are free to switch between languages when they want. Indeed, several studies have reported that language switching occurs naturally both in conversations as well as within sentences (e.g., Milroy & Muysken, 1995; Myers-Scotton & Lake, 1995). In the current study, we examine when and why bilinguals switch between languages when they are free to name pictures in their language of choice. In this context, we measure whether there are differences in the amount of language control needed during cued and voluntary switching.

Language switching paradigms

Many studies have examined language switching through the use of a cued picture or digit naming paradigm (e.g., Costa & Santesteban, 2004; Meuter & Allport, 1999). In these paradigms, participants are asked to name items in a *blocked* and a *mixed* context. Within the blocked context, all items must be named in one pre-specified language. The mixed context, in contrast, requires participants to switch between languages according to a cue. This produces switch trials (the current trial's language differs from the previous one) as well as non-switch trials (two consecutive trials are named in the same language). Bilinguals usually take longer to respond on switch than on non-switch trials (the 'switching cost'), reflecting the effort associated with system reconfiguration. Inhibitory control may be an important mechanism underlying language switching through inhibition of the non-target language and subsequent (re)activation of the previously inhibited representations (Green, 1998). These switching costs reflect a temporally local type of language control at the trial level. When non-switch trials from the mixed condition are compared to trials in the blocked condition, response times (RT) are typically longer in the mixed condition (the 'mixing cost'). This mixing cost is taken as an indication that more global, proactive control mechanisms are needed to maintain two languages or tasks (cf. Rubin & Meiran, 2005), and that such maintenance requires cognitive resources.

Voluntary language switching

In daily life, bilinguals need to monitor cues (e.g., the language(s) spoken by the interlocutor) in order to switch between languages when needed. However, in some circumstances, bilinguals can freely switch between languages. In Green and Abutalebi's (2013) Adaptive Control Hypothesis, three bilingual language contexts are described that are associated with different levels of language control. In the single language context, bilinguals use their two languages in strictly separated settings (e.g., one language at home, one language at work). This type of language use mainly places demands on global control mechanisms such as goal maintenance and conflict monitoring. In the dual-language context, both languages can be used in the same setting but with different interlocutors (e.g., both languages are used at work, but one language with person A and the other language with person B). In this setting, language switching occurs frequently but not with the same interlocutor. This type of language context requires constant monitoring of the circumstances in order to select the appropriate language and as such is argued to require a relatively high level of language control. In contrast, in the third context ('dense code-switching'), less control is needed. In this context, a bilingual speaker is surrounded by other bilinguals who speak the same languages, allowing them to switch freely, even within conversations. According to the Adaptive Control Hypothesis, this switching context does not place additional demands on cognitive processes such as goal maintenance, cue detection, and response inhibition. In contrast, the dual-language context, which is most similar to the typical cue-based laboratory measurements of language switching, does increase the need for these cognitive processes. Thus, the switching and mixing costs that have been linked to language switching may be related to the use of a cue-based switching paradigm that requires a higher level of control. If lower levels of control are needed in a voluntary context, these switching and mixing costs may be smaller or even absent.

Despite voluntary language switching frequently occurring in daily life in many bilingual societies, relatively little experimental research has examined when and why bilinguals switch between languages and how voluntary switching may affect the costs observed in cued paradigms. While cued language switching explicitly requires top-down control (as switches are driven by explicit instructions), it is unclear whether voluntary switching also requires top-down control. Instead (or in addition), voluntary switching may be driven by bottom-up processes (i.e., lexical access) if bilinguals just use the language that comes to mind first.

Gollan and Ferreira (2009) examined voluntary language switching in Spanish-English bilinguals and showed that participants switched frequently. They also found that unbalanced bilinguals on average switched less often (24% of the trials) than balanced bilinguals (35% of the trials). A switching cost was observed, suggesting that switching remained costly even though done voluntarily[1]. The mixing costs showed a more complex pattern, with mixing *costs* for balanced bilinguals and mixing *benefits* for unbalanced bilinguals when using the non-dominant language. Gollan and Ferreira linked this mixing benefit to unbalanced bilinguals monitoring the accessibility of items in the weaker language and only choosing this weaker language when an item is easily accessible. As a consequence, the mixing benefit in this group of bilinguals may be interpreted as reflecting the avoidance of naming less accessible items in the voluntary condition, something that cannot be avoided in the blocked condition. However, further analyses in Gollan and Ferreira (2009, Experiment 1) showed that a mixing benefit was not only found for items that were always named in the weaker language, but also for items that did not show such a strong language preference. This suggests that lexical access cannot fully explain the mixing benefit.

A study on voluntary language switching in children (Gross & Kaushanskaya, 2015) provides further evidence that language choice is related to lexical access. Items

[1] Note, however, that Gollan and Ferreira's (2009) Figure 2 suggests that this switching cost was not present in balanced bilinguals.

were more frequently named in the non-dominant language when they had a higher frequency of use, were more likely to be acquired early in life in the non-dominant language, and were less likely to have alternative naming options in the non-dominant language. Furthermore, as in Gollan and Ferreira (2009), different mixing effects were observed for the dominant and non-dominant language. A mixing cost was found in reaction times for the dominant but not for the non-dominant language. This asymmetry was also present in the accuracy scores, but here a mixing benefit was found for the non-dominant language while no effect was observed for the dominant language. However, the presence of switching costs suggested that voluntary switching remained costly.

While most studies on voluntary language switching have reported costs associated with switching, Blanco-Elorrieta and Pylkkänen (2017) demonstrated that voluntary switching can be cost-free. Participants named pictures in response to artificial cues (colours), faces of monolinguals ('monolingual cued', requiring them to use a specific language), and faces of bilinguals ('bilingual voluntary'). The authors observed behavioural switching costs in response to artificial cues, but not when bilinguals responded to natural cues (i.e., the monolingual cued or bilingual voluntary conditions). Furthermore, when bilinguals had to switch in response to artificial cues, increased activation was observed in the anterior cingulate cortex (ACC) and dorsolateral prefrontal cortex (dlPFC), two areas that have been linked to language switching and executive control. In the monolingual cued condition using natural cues, this increased switch activation was found in the earliest time window of analysis only. For bilingual voluntary trials, this increased activation was not found at all. Thus, language switching was argued to be most effortful in response to artificial cues and not costly at the behavioural or neural level when done voluntarily.

Lexical access and top-down control

Taken together, studies investigating voluntary language switching suggest that mechanisms related to inhibitory/executive control as well as lexical access may be involved. Language choice appears to be at least partly driven by lexical access because easier items are often named in the non-dominant language while the dominant language is used for more difficult items. Furthermore, mixing benefits rather than costs have been observed for the non-dominant language of unbalanced bilinguals, which could be due to naming easier items in the non-dominant language. However, it has been suggested that the mixing benefit cannot fully be explained by lexical access (Gollan & Ferreira, 2009) and might reflect a more general benefit of voluntary language mixing.

The presence of switching costs even in voluntary switching tasks suggests that top-down control processes are involved. However, not all studies have observed switching costs (cf. Blanco-Elorrieta and Pylkkänen, 2017). Others have suggested that switching costs may depend on the extent to which voluntary language switching is bottom-up and driven by lexical access (Kleinman & Gollan, 2016). Switching costs were absent when participants were instructed to choose the easiest language for each picture and then use that language to name that picture for the rest of the task. In contrast, when participants were given the instruction to just choose the language that came to mind first for each item, without having to use that language for the rest of the task, switching costs were similar for the voluntary and cued switching tasks. Furthermore, Gollan, Kleinman, and Wierenga (2014) noted that switching costs in voluntary tasks may be affected by item repetition. When pictures were not repeated, switching costs were similar for cued and voluntary language switching. However, when pictures were repeated, as is often done in language switching studies, switching costs were smaller for voluntary than cued switching.

Current study

The current study aimed to examine in more detail when and why bilinguals voluntarily switch between their languages and how lexical access as well as more top-down control processes are involved.

In order to assess voluntary language switching in highly proficient bilinguals who are used to language switching in daily life, we collected data from bilingual Spanish-Basque young adults living in the Basque Country, a region in the north of Spain. This region is characterised by a large number of Spanish-Basque bilinguals, many of whom acquired both languages from birth and have reached a high proficiency level in both. The two languages are omnipresent in the daily society, allowing bilinguals to freely switch between both languages.

To examine the amount of language control needed in cued versus voluntary switching, we compared both tasks across two experiments. In Experiment 1, bilinguals completed a picture naming task asking them to use one language only ('blocked condition') or voluntarily choose one of the languages ('mixed condition'). In Experiment 2, a different group of bilinguals in addition completed a cued picture naming task. This allowed us to directly compare the mixing and switching costs in the cued and voluntary tasks within the same group of participants.

A joint analysis of the data from Experiment 1 and 2 was conducted to examine when and why bilinguals switch. Our main goal was to clarify the relationship between lexical access and language choice that has been proposed in previous studies. To do so, we used the RTs from the first blocked condition, in which participants had to name all pictures in only one language, to estimate the speed with which each participant named each item upon its first presentation in Basque and Spanish. We then examined whether this naming speed was related to language choice in the voluntary mixed condition.

In addition to the picture naming task, participants also completed a range of tasks assessing language proficiency, language use, inhibitory control, and working memory performance. This allowed us to examine whether voluntary language switching is related to language-based factors such as language proficiency as well as to general cognitive control-related factors such as working memory or inhibitory control.

Experiment 1

Methods

Participants

Experiment 1 was completed by 55 Spanish-Basque bilinguals. All participants had normal or corrected-to-normal vision, no known neurological, reading, or hearing impairments, and gave informed consent. The study was approved by the BCBL Ethics Review Board and complied with the guidelines of the Helsinki Declaration. One participant's data was not used as she reported English to be her first language. The remaining 54 participants (33 female) were on average 22.3 years old (*SD* = 2.5, range 18-30) and reported 18.0 years of education (*SD* = 3.2). Most (42 participants) had completed their primary and secondary education in Basque; 11 had received their education in a mixed language model with classes taught in Spanish and Basque; and one participant had received full education in Spanish with Basque as a subject. All participants were highly proficient in both Spanish and Basque and acquired both languages before or at the age of 7. However, in order to assess possible effects of language proficiency on voluntary language switching performance, we included a range of different proficiency profiles ranging from balanced bilinguals to Spanish-dominant bilinguals.

In order to formally assess language proficiency, use, and exposure, participants completed a series of objective and subjective proficiency measurements. These included an interview, a picture naming task, and a lexical decision task (LexTALE). In addition, they provided a range of self-ratings with respect to language proficiency, exposure, and use in Spanish and Basque that form part of the BEST proficiency test (de Bruin, Carreiras, & Duñabeitia, 2017). In the five-minute interview, participants were asked several questions ranging in difficulty. Their proficiency and fluency were rated on a scale from 1 ('lowest level') to 5 ('native or native-like level') by native speakers of Basque and Spanish. In the picture naming task, participants had to name 65 non-cognate pictures in Basque and Spanish. The LexTALE is a short computerised lexical decision task, consisting of 60 words and 30 non-words in the Spanish version and 50 words and 25 non-words in the Basque version. Lastly, participants completed a questionnaire with various questions about their proficiency in and use of their languages. Participants were asked to self-rate on a scale from 0 to 10 their proficiency in each language in terms of speaking, understanding, reading, and writing, as well as a general proficiency score. In terms of exposure, participants were asked to rate on a scale from 0 to 100% how often they were exposed to each language. As a measurement of active language use, we also report the participants' ratings of how often they speak each language on a scale from 0 to 100%. The objective proficiency measurements and questionnaire were completed by the participants when they signed up for the participant database.

In addition, participants were asked to complete two verbal fluency tasks in separate blocks of Basque and Spanish. In the letter fluency task, they were asked to name words starting with an 'S' within one minute. In the category fluency task, they named as many animals as possible within one minute. The language order and task order were counterbalanced across participants. The average scores, standard deviations, and ranges of all language measurements are provided in Table 1. In an additional questionnaire, participants provided self-ratings of daily language switching behaviour, indicating on a scale from 1 ('never') to 4 ('all the time') how often they a) switch on a daily basis, b) switch within a conversation, and c) switch within a sentence. On average, participants reported frequent daily language switching (*M* = 2.6, *SD* = 1.1) and occasional switching within conversations (*M* = 2.1, *SD* = 0.8) and sentences (*M* = 1.8, *SD* = 0.7). In addition to Basque and Spanish, most participants reported having acquired English as a third language (average English picture naming score (scale 0-65): *M* = 44.7, *SD* = 9.8).

Table 1. *Summary of objective and subjective measurements of language proficiency, exposure, and use of Spanish (left) and Basque (right) for Experiment 1. There was a significant difference between Spanish and Basque on all measurements.*

	Spanish			**Basque**		
	Mean	SD	Range	Mean	SD	Range
Age of Acquisition	0.2	0.6	0-3	1.9	1.7	0-7
Picture naming (0-65)	64.6	0.7	62-65	55.2	7.4	42-65
LexTale[2] (0-100%)	94.1	5.1	76-100	87.7	7.2	67-99
Interview (1-5)	5.0	0.0	5-5	4.3	0.6	3-5
Self-rated proficiency (0-10)[3]						
Speaking	9.6	0.6	8-10	8.2	1.5	4-10
Understanding	9.7	0.5	8-10	9.0	1.0	6-10

[2] The LexTale was completed in both languages by 52 participants.
[3] Self-rated proficiency was completed for both languages by 49 participants.

Writing	9.2	1.1	6-10	8.3	1.6	4-10
Reading	9.6	0.8	7-10	9.0	1.1	6-10
General	9.3	0.9	7-10	8.3	1.4	4-10
%exposure						
(0-100)	56.5	13.8	30-80	31.9	13.9	10-60
%speaking						
(0-100)	62.0	17.1	20-90	30.4	16.4	10-80
Verbal fluency						
Animals	24.7	7.0	13-43	17.0	5.1	10-29
Letters	12.8	4.9	4-24	6.8	3.2	1-14

Procedure

All participants completed the voluntary language switching task as well as several background measurements within one session lasting about 1 to 1.5 hours. The session always started with the voluntary language switching task, followed by a reverse digit span, backward Corsi block tapping, verbal Stroop, and numerical Stroop task in counterbalanced order (see below). Lastly, participants completed a short questionnaire and the verbal fluency tasks.

Voluntary language switching task.

In the voluntary language switching task, participants were asked to name pictures in Spanish and/or Basque. As detailed below, there were multiple parts in the voluntary language switching task, parts in which all naming was done in one language (i.e., either Spanish or Basque), and a part in which the languages were mixed. All instructions were provided on the computer screen in both Basque and Spanish, except for the instructions for the blocked conditions which were only given in the language that had to be used. For the instructions given in both languages, the order in which the languages appeared on the screen (top or bottom half) was counterbalanced across participants. Prior to the task, participants were familiarised with the stimuli by showing each picture with the corresponding words in Basque and Spanish one by one on the screen. This familiarisation was included to minimise the number of errors due to not recognising the pictures or using the wrong word (e.g., 'horse' instead of 'donkey').

The task included three sections, always starting with a blocked condition, followed by the mixed condition, and ending with another blocked condition. Each blocked condition included two subsections. In each subsection, participants were asked to name all pictures in Basque only or in Spanish only. Half of the participants completed the first blocked condition in Basque first and Spanish second, with the order reversed in the second blocked condition; the other half followed the reverse language order. Each blocked condition consisted of 60 items (30 per language). In the mixed condition, participants were free to decide in which language they wanted to name each picture. The following instructions were provided prior to the mixed condition: 'In the following part, you can name the pictures in Spanish or Basque. You are free to switch between languages whenever you want. Try to use the word that comes to mind first, but don't use the same language throughout the whole task'. The mixed condition consisted of 360 trials, distributed across 6 blocks. A relatively large number of mixed trials was used to get a stable estimate of switching frequency. The first blocked condition and the mixed condition were preceded by practice trials (four trials per language in the blocked condition and eight trials in the mixed condition; the practice trials showed pictures that were not used in the experiment).

Each trial started with the presentation of a fixation cross for 500 ms, followed by the presentation of the picture that had to be named. The pictures stayed on the screen for 2500 ms, regardless of when a response was given. Participants were asked to name the pictures aloud while avoiding hesitations and their responses were recorded.

Thirty individual pictures were used that were repeated four times in the blocked conditions and twelve times in the mixed condition (see appendix A for stimulus materials). The images were selected from the MultiPic database (Duñabeitia et al., 2018) that provides norms in six languages, including Spanish. Only images with a naming agreement above 70% in Spanish were included in the initial selection. In order to assess Basque responses, we asked four native Basque speakers to name the pictures. Based on these responses, we only included items that were named the same way by the four native speakers. In addition, we asked a Basque native speaker to exclude those items that have a Basque translation but for which native speakers typically use the Spanish word. In order to exclude identical cognates, we then calculated the Levenshtein distance (relative to word length) and excluded words that reached a similarity score above .40 as well as words that were marked as cognates by the native speaker. As most participants reported having some knowledge of English, we also excluded English cognates. The chosen picture names were matched on Spanish and Basque word frequency, number of phonemes, and number of syllables (see Appendix A).

The task was presented using Psychopy 1.83.04 (Peirce, 2007). Stimuli were 200x200 pixels and were presented using a white background on a Viewsonic E90f monitor, with 90Hz refresh rate, and a screen resolution of 1024x768. Responses were recorded through a Sennheiser PC 151 headset with microphone.

Working memory and inhibitory control tasks.

All background tasks were presented in Spanish in Experiment Builder (SR Research). Working memory span was assessed through a reverse digit span and backward Corsi block tapping task. In the digit span task, participants listened to series of digits that they were asked to recall verbally in reverse order. The task started with a series of two digits, with the span being increased by one on every two trials. The maximum length that could be reached was 8 digits (14 trials in total). If participants answered both trials of a certain span incorrectly, the task was ended. In the Corsi task, a similar approach was used, but in the visual domain. Participants were asked to remember the sequence in which shapes lit up on the screen and to repeat this order backwards by pointing at the shapes afterwards. For both the digit and Corsi tasks, the total number of correct trials was taken as an indication of working memory span.

Verbal and numerical Stroop tasks were used as measurements of verbal and non-verbal inhibitory control (cf. Antón, Fernández García, Carreiras, & Duñabeitia, 2016). In the verbal Stroop task, participants were presented with colour names and were asked to name the visual colour of the word ('rojo' ('red'), 'azul' ('blue'), or 'amarillo' ('yellow')). In the congruent condition (24 trials), the colour name matched the colour in which the word was presented (e.g., 'rojo' presented in red). In the incongruent condition (24 trials), the colour name mismatched the visual colour (e.g., 'rojo' presented in blue). In the neutral condition (24 trials), three non-colour words ('ropa' ('clothing'), 'avión' ('airplane'), 'apellido' ('surname')) were presented in a colour. In the non-verbal Stroop task, a number-size paradigm was used in which participants saw two digits, one on the left and one on the right side of the screen. Participants were asked to indicate with a button press whether the digit on the left or right side was larger in font size. On congruent trials (24 trials), the larger number was presented in a larger font (e.g., bigger 7 with smaller 1). On incongruent trials (24 trials), there was a mismatch between numeric size and font size (e.g., smaller 7 with bigger 1). On the neutral trials (24 trials), the same number was presented in different sizes (e.g., bigger and smaller 1). Following previous studies observing a link between language switching costs and inhibition costs (e.g., de Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014; Linck, Schwieter, & Sunderman, 2012), we took the RT difference between incongruent and congruent trials ('Stroop cost') as a measurement of inhibitory control. To improve split-half reliability, the Stroop cost was calculated after outlier removal (2.5 *SD* above or below the mean per trial type) [4].

Data analysis

The data are available at: osf.io/drth2.

During the picture naming task, answers were scored by the research assistants running the experiment. The response language was also coded during the task and trials were coded as switch or non-switch trials afterwards. Response accuracy was scored as follows: A) no or late response; B) correct response; C) correct language but wrong word (e.g., 'horse' instead of 'donkey'); D) wrong language (blocked condition only); E) hesitation; F) combination of both languages (e.g, 'cab[allo]-horse'); G) items

[4] Other studies have taken the difference between incongruent and *neutral* trials instead, which could be argued to be a clearer reflection of inhibition. However, split-half reliability analyses across the two experiments (odd versus even trials) showed a lower reliability of this measurement in the current study (Spearman-Brown coefficient .54 for verbal Stroop and .43 in the non-verbal task). For the score calculated as the difference between incongruent and congruent trials, which was used here, split-half reliability coefficients were .65 for the verbal task and .80 for the non-verbal task.

in the mixed condition that could not be identified as switch or non-switch trials (i.e., the first trial after a break and trials preceded by trials scored as A or F); H) recording failures (5 trials in Experiment 1 and 13 in Experiment 2 that were scored as being correct during the experiment). Responses were recorded and response times were derived using Chronset software (Roux, Armstrong, & Carreiras, 2017) and checked with CheckVocal (Protopapas, 2007). Response times for answers with hesitations were manually corrected (i.e., in the case of 'Uh…donkey', the response onset would be measured at the /d/) and counted as accurate responses.

Switching frequency in the voluntary mixed condition was calculated as the number of switch trials divided by the total number of trials. For this count, the correct trial types B, E, and H were included as well as inaccurate answers of type C (wrong word). We excluded trials of type G as they could not be classified as switch or non-switch trials. For all further analyses, inaccurate responses (type A, C, D, F) were excluded (less than 5% of trials for all conditions) as well as trials of type G and H.

All analyses used the following approach. Data were analysed with mixed-effect models in R using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). RTs were log transformed and analysed with linear mixed-effect models. To reduce collinearity, all continuous fixed effects were z-scored and the two-level categorical predictors were coded as -0.5 and 0.5. Participants and items were included as random effects.

We always started with a full model including all fixed effects of interest, random intercepts for participants and items and slopes for all predictors (i.e., a maximal structure, Barr, Levy, Scheepers, & Tily, 2013). When models did not converge, the following approach was used on the random effects structure: We first removed correlations between the random slopes and the random intercepts (‘no random correlations’, Barr et al., 2013). If the resulting model still did not converge, we built down the item structure as this is expected to explain the least variance, thus keeping a full slope structure for participants. We removed item slopes that explained the least variance until the model reached convergence.

We report the full models including all fixed effects, even the non-significant ones. Keeping the random effects structure constant, we then also selected the best fitting model through a stepwise procedure. First, we removed all non-significant predictors and compared the simpler model to the full model using likelihood-ratio chi-square tests (Baayen, 2008). When the simpler model excluding non-significant predictors did not differ significantly from the more complex model, the simpler model with fewer predictors was preferred. Then, we removed each significant predictor from the simpler model and examined whether this worsened the model. As these model comparisons involved different fixed effects, we used parameters estimated with Maximum Likelihood (ML) rather than Restricted Maximum Likelihood (REML). However, we report the values from the models estimated with REML.

The models were checked for collinearity between predictors through VIF.mer (Frank, 2011). In all models, VIFs were below 2.5. *T* and *z* values > 2 were interpreted as significant (Gelman & Hill, 2007).

Analyses were conducted across all data points but by-participant means are provided in the text and figures. For the RTs, the reported averages are based on raw values even though the analyses were performed on log RTs.

For Experiment 1, one model was constructed to examine mixing costs and one to examine switching costs. The model examining mixing costs only included non-switch and blocked trials; the model examining switching costs included non-switch and switch trials. Both models were constructed with log RTs as the dependent variable (DV) and language, trial type, and language x trial type as fixed effects. For the factor language, Basque was coded as -0.5 and Spanish as 0.5. For the model on mixing costs, trial type was contrast coded with blocked trials coded as -0.5 and non-switch trials as 0.5. For the model on switching costs, coding was -0.5 for non-switch trials and 0.5 for switch trials. The maximal random effects structure was used (including intercepts for subjects and items as well as slopes for language + trial type + language x trial type for subjects and items).

Results

An initial examination of the switching frequency showed that participants on average switched on 40.8% (*SD* = 11.0) of the trials. Over 80% of participants switched between 30% and 50% of the time (see Figure 1, left panel) with the switching percentage ranging from 8% to 77%. Across the items that could be classified as switch or non-switch trials, 56.3% were named in Basque (*SD* = 11.3, range 22%-96%), of which 38.0% (*SD* = 13.0) were switch trials. Of the items named in Spanish, 49.8% (*SD* = 14.6) were switch trials. One participant who switched on 8% of the trials named almost all items in Basque and always switched back to Basque after naming an item in Spanish. As he therefore did not produce any non-switch trials in Spanish, he was excluded from further analyses.

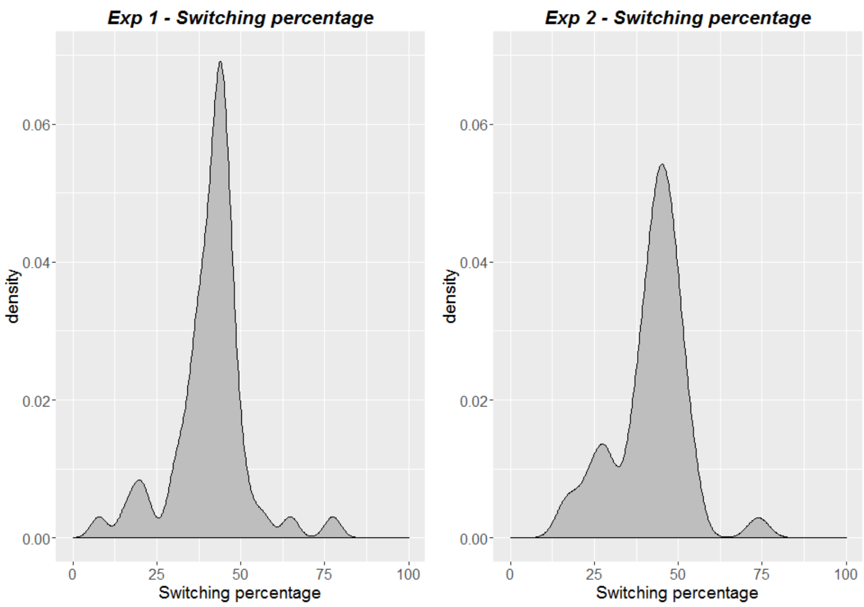

Figure 1. Density plot showing the distribution of switching percentage across participants in the voluntary language switching task for Experiment 1 (left) and Experiment 2 (right). The area under the curve between point A and B (e.g., 25% - 50%) reflects the probability of a value falling between those points A and B (with the total area under the curve being 1).

Average accuracy was high in both the Basque and Spanish blocked conditions (Blocked Basque *M* = 95.1, *SD* = 5.7; Blocked Spanish *M* = 96.1, *SD* = 3.8) and close to ceiling in the voluntary condition (overall *M* = 99.1, *SD* = 1.4; of the items that could be classified as being Basque: *M* = 99.2, *SD* = 1.5; Spanish *M* = 99.5, *SD* = 1.3). As accuracy was close to ceiling and not of specific interest for the current study, it was not analysed further.

In our analysis of reaction times, RTs more than 2.5 *SD* above or below the mean (calculated on the log RTs per participant per trial type and language, 2.2% of all correct trials) were removed.

In the model examining mixing costs, there was a main effect of trial type (β = -0.066, *SE* = 0.014, *t* = -4.746), with blocked trials (*M* = 876.2, *SD* = 98.9) being significantly slower than non-switch trials (*M* = 806.6, *SD* = 112.4), reflecting a mixing benefit (M = -69.6, *SD* = 93.1; see Figure 2). There was a main effect of language (β = 0.056, *SE* = 0.015, *t* = 3.874), reflecting Basque trials (*M* = 812.1, *SD* = 103.2) being faster than Spanish trials (*M* = 868.2, *SD* = 113.7). Language did not interact significantly with the mixing benefit (β = -0.008, *SE* = 0.011, *t* = -0.695).

The model on switching costs showed that switch trials (*M* = 852.6, *SD* = 124.3) were slower than non-switch trials (*M* = 806.6, *SD* = 112.4; β = 0.042, *SE* = 0.007, *t* = 6.398), reflecting a switching cost (*M* = 46.0, *SD* = 43.4; see Figure 2). There was a main effect of language (β = 0.049, *SE* = 0.010, *t* = 4.889) but no interaction with switching costs (β = -0.005, *SE* = 0.011, *t* = -0.443).

For both models, the best model was the simple model including language and trial type, with no interaction between the two. Removal of either language or trial type resulted in a significantly worse model.

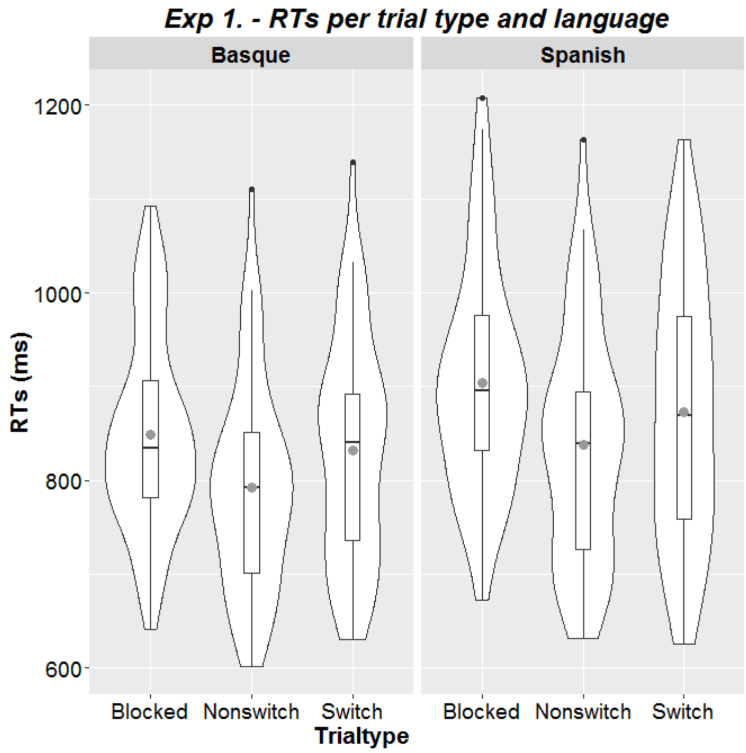

Figure 2. Violin plots showing the distribution of the RTs for the voluntary task for each language (left: Basque; right: Spanish) for each trial type (blocked, non-switch, and switch trials). The plot outline shows the density of data points at the different RTs. The boxplot shows the interquartile range, the thin vertical black line the 95% confidence interval. The median is indicated by the thin horizontal black line, while the grey dot represents the mean.

Discussion

The reaction time analysis in Experiment 1 showed a mixing benefit for both languages when bilinguals could voluntarily use their languages. This is in contrast to the mixing costs that are typically observed in cued switching tasks. The voluntary switching task did show a significant switching cost. The mixing benefit and switching cost were similar for Basque and Spanish.

Basque was the preferred and faster language in this experiment, which is surprising given that participants either had balanced proficiency in both languages or were more proficient in Spanish than Basque. Furthermore, Basque and Spanish items were matched on overall frequency and word length, and accuracy was similar for the two languages, suggesting that the pictures were not necessarily easier to name in

Basque than Spanish. Faster RTs in Basque than Spanish and a preference for Basque could reflect reversed dominance effects that have been observed previously with responses being faster in the non-dominant than dominant language (cf. Kleinman & Gollan, 2018, for a discussion). These reversed dominance effects may be caused by greater inhibition of the dominant language, leading to relatively slow responses in that language and relatively faster responses in the weaker language. However, these reversed dominance effects have typically been observed in mixed language conditions only, while the current study showed faster Basque responses on the blocked trials too, even the first time the pictures were named. Several socio-linguistic factors could also explain this language preference, although all are speculative. The great majority of participants received their schooling in Basque, so Basque could have been the more natural language to use in a picture-naming task. Secondly, while all instructions were provided on the screen in both languages, the experimenters were instructed to only use one language with the participants (i.e., the language in which the participant initiated the conversation) and consequently used Basque with most participants. This could have primed the participants' overall language preference.

As Experiment 1 did not include a cued switching task, we do not know whether the voluntary switching cost observed here was comparable to or smaller than cued language switching. In Experiment 2, another group of Spanish-Basque bilinguals therefore completed a cued switching task in addition to the voluntary switching task. In this way, we could directly compare switching costs across tasks. In addition, this allowed us to examine whether the same group of bilinguals would show a mixing benefit in a voluntary picture naming task but a mixing cost when being told which language to use.

Experiment 2

Methods

Forty-five participants took part in Experiment 2. Vocal responses from two participants were not recorded and as such could not be included in the analysis. The average age of the remaining 43 participants (27 female) was 25.7 years (*SD* = 4.4, range 18-35) and participants had completed 16.7 years of education (*SD* = 2.1). Thirty-five participants had received Basque education and the remaining eight completed their education in Basque and Spanish. The language background from these participants was similar to those tested in Experiment 1 (see Table 2 for the results from the objective language proficiency measurements and questionnaires). On a scale from 1 to 4, participants rated their language switching frequency as *M* = 3.0 (*SD* = 1.0) on a daily basis, *M* = 2.3 (*SD* = 1.0) within a conversation, and *M* = 2.0 (*SD* = 0.8) within sentences. For most participants, English was the third language (*M* picture naming (0-65): 47.5, *SD* = 10.8).

Table 2. *Summary of objective and subjective measurements of language proficiency, exposure, and use of Spanish (left) and Basque (right) for Experiment 2. There was a significant difference between Spanish and Basque on all measurements.*

	Spanish			**Basque**		
	Mean	SD	Range	Mean	SD	Range
Age of Acquisition	0.5	1.1	0-3	1.3	1.6	0-6
Picture naming (0-65)	64.7	0.6	62-65	58.9	6.4	41-65
LexTale[5]						
(0-100%)	94.5	3.5	83-100	92.1	4.7	82-100
Interview (1-5)	5.0	0.0	5-5	4.6	0.7	3-5
Self-rated proficiency						
(0-10)[6]						
Speaking	9.4	0.9	7-10	8.6	1.5	5-10
Understanding	9.5	0.8	7-10	9.1	1.0	6-10
Writing	8.8	1.2	6-10	8.4	1.3	5-10
Reading	9.5	0.8	7-10	8.9	1.4	5-10
General	9.5	0.7	7-10	8.8	1.3	5-10
%exposure						
(0-100)	51.6	19.0	10-90	34.4	18.4	10-90
%speaking						
(0-100)	53.5	23.5	10-100	37.7	23.3	10-80
Verbal fluency						

[5] The LexTale was completed in both languages by 39 participants.
[6] Self-rated proficiency was completed for both languages by 41 participants.

Animals	25.3	4.9	13-37	18.5	5.1	8-31
Letters	12.4	4.0	2-20	7.8	3.1	2-15

Experiment 2 consisted of two sessions. One session was identical to the one described for Experiment 1, consisting of the voluntary picture naming task, inhibitory control tasks, working memory tasks, and several background measurements. Another session was conducted a few days before or after this session, during which participants completed a cued picture naming task[7].

The cued picture naming task consisted of a different set of thirty pictures (see Appendix B) that were selected according to the procedure described for the voluntary task in Experiment 1. The items used in the cued task were very similar to the ones used in the voluntary task in terms of number of phonemes, syllables, and log frequency in Basque and Spanish (see Appendices A and B for more details). The cued switching task followed the procedure of the voluntary switching task, except that there were cues indicating the language in which each item had to be named. Within the mixed condition, half of the trials were switch trials and half of the trials had to be named in each language. For each participant, the items were pseudo-randomised across trials such that items were never repeated. Across the participants, all items were required to be named approximately half of the time in Basque and half of the time in Spanish (range across items: 45-57%). The mixed condition consisted of a predictable and an unpredictable part that were included for a study examining age effects on language switching. In the predictable condition, the cues followed the order Basque-Basque-Spanish-Spanish. In the unpredictable condition, the order of cues was random. For the purpose of the current study, the analyses are collapsed across these two parts.

Each picture was preceded by the Spanish or Basque flag, which remained on the screen for 500 ms and was then presented in a smaller format above the actual picture. The size of the cue was 200x100 pixels when presented alone and 100x50 pixels when presented above the stimulus. To ensure that a switch between cues took place even on non-switch trials (and thus to avoid a confound between cue and

[7] Of the 43 participants, 15 completed the voluntary task on day 1 and the cued task on day 2. Twenty-eight participants completed the cued task on day 1 and the voluntary task on day 2.

language switching), we used two versions of the Spanish flag and two versions of the Basque flag. Only using one cue per language would mean that switching costs not only reflect language switching but also cue switching (e.g., Logan & Bundesen, 2003; Mayr & Kliegl, 2003). As such, with one cue per task, larger switching costs on the cued than voluntary task could reflect cue switching costs that would only have been present in the cued task. Therefore, using two cues per language allowed for a closer comparison of the actual language switching costs in both tasks. In order to minimise differences between the blocked and mixed condition, the cues were also presented during the blocked conditions even though they were redundant in that context.

Data analysis

In order to examine mixing and switching costs across the cued and voluntary conditions, models of mixing costs and of switching costs were constructed (one including blocked and non-switch trials for mixing costs, and one including non-switch and switch trials for switching costs) with log RTs as the DV and task, trial type, language, and task order as well as all interactions as predictors. We included task order (cued or voluntary task first) because previous studies (e.g., Kleinman & Gollan, 2016) have observed order effects. Task was coded as -0.5 for cued and 0.5 for voluntary. Trial type and language were coded in the same way as in Experiment 1. Task order was manipulated between-subject and task between-item, so the random effects structure included task order slopes by item only and task slopes by subject only. The models with the maximal random effects structure did not converge. After removal of random correlations, the models converged with all subject and item slopes.

Results

Average switching frequency was similar to Experiment 1 (*M* = 41.9%, *SD* = 10.6), ranging from 16% to 74% (see Figure 1, right panel). Over two-thirds of the participants switched on between 30% and 50% of the trials. In terms of language choice, 57.0% of the items classifiable as switch or non-switch trials were named in Basque (*SD* = 8.2, range 42%-82%), of which 38.0% (*SD* = 11.8) were switch trials. Of the items named in Spanish, 49.6% (*SD* = 12.4) were switch trials.

As in Experiment 1, accuracy was very high in all conditions and was not analysed further. In the voluntary blocked condition, mean accuracy was 97.7% (*SD* = 2.7) in Basque and 97.9% (*SD* = 3.1) in Spanish. Accuracy was close to ceiling in the voluntary mixed condition (*M* = 99.2, *SD* = 1.3; items classified as Basque *M* = 99.3%, *SD* = 1.4; items classified as Spanish *M* = 99.6%, *SD* = 0.8). In the cued blocked condition, average Basque accuracy was 96.1% (*SD* =5.6) and Spanish accuracy was 97.9% (*SD* = 2.7). In the cued mixed condition, average accuracy was 97.8% (*SD* = 2.8) for Basque and 97.6% (*SD* = 3.0) for Spanish.

RTs more than 2.5 *SD* above or below the mean (calculated on log RTs per participant and per language, task, and trial type; 2.1% of all correct trials in the voluntary task and 2.0% in the cued task) were removed for the RT analysis. As different stimuli were used for the voluntary and cued task, we assessed whether there were any baseline differences in the blocked condition. There was no significant difference ($F(1,42) = 0.785$, $p = 0.381$) between cued blocked RTs (*M* = 841.4, *SD* = 98.3) and voluntary blocked RTs (*M* = 851.3, *SD* = 104.1), nor a significant interaction between task and language ($F(1,42) = 0.468$, $p = 0.498$; Cued Basque *M* = 807.0, *SD* = 107.6; Voluntary Basque *M* = 822.4, *SD* = 105.1; Cued Spanish *M* = 875.5, *SD* = 105.3; Voluntary Spanish *M* = 880.6, *SD* = 119.1).

Figure 3 presents the RT results for the two tasks (cued, voluntary), broken down by language (Basque, Spanish) and by trial type (blocked, non-switch, switch). The model of mixing costs showed no main effect of task (β = -0.018, *SE* = 0.020, *t* = -0.907). There was a main effect of language (β = 0.070, *SE* = 0.012, *t* = 5.716), reflecting the fact that Basque responses were faster than Spanish (see Figure 3). There was no main effect of trial type (β = 0.005, *SE* = 0.009, *t* = 0.574), but trial type interacted with task (β = -0.063, *SE* = 0.017, *t* = -3.726). Blocked trials were faster than non-switch trials in the cued task (mixing cost *M* = 25.3, *SD* = 55.4) but slower than non-switch trials in the voluntary task (mixing benefit *M* = -34.3, *SD* = 75.1; see Figure 3). Language did not interact with mixing costs (β = -0.010, *SE* = 0.009, *t* = -1.106) or task (β = -0.014, *SE* = 0.016, *t* = -0.837), nor was there a three-way interaction (β = 0.017, *SE* = 0.020, *t* = 0.849). There was no main effect of session order (*t* = -1.118) nor were there any significant interactions with session order (all *t*s < 1). The best model was the model with a main effect of language and an interaction between mixing costs/benefits and task. Removal of any of these effects resulted in a significantly worse model.

The model of switching costs showed a significant effect of task (β = -0.051, *SE* = 0.022, *t* = -2.292). Responses were faster in the voluntary mixed condition (*M* = 832.1, *SD* = 118.3) than the cued mixed condition (*M* = 882.7, *SD* = 114.4). There was a main effect of language (β = 0.059, *SE* = 0.012, *t* =5.097) and trial type (β =0.034, *SE* = 0.004, *t* = 7.576). Trial type did not interact with task (β = -0.003, *SE* = 0.007, *t* = -0.369), reflecting the similar switching costs for the cued (*M* = 32.3, *SD* = 37.5) and voluntary (*M* = 39.7, *SD* = 39.7) tasks (see Figure 3). Language did not interact with switching costs (β = -0.012, *SE* = 0.008, *t* = -1.515) or task (β = 0.004, *SE* = 0.016, *t* =0.251) nor was there a significant three-way interaction (β = 0.022, *SE* = 0.013, *t* = 1.663). There was no main effect of or an interaction with session order (*t*s < 1). The best model was the model including main effects of language, task, and trial type; removal of any of these effects led to a significantly worse model.

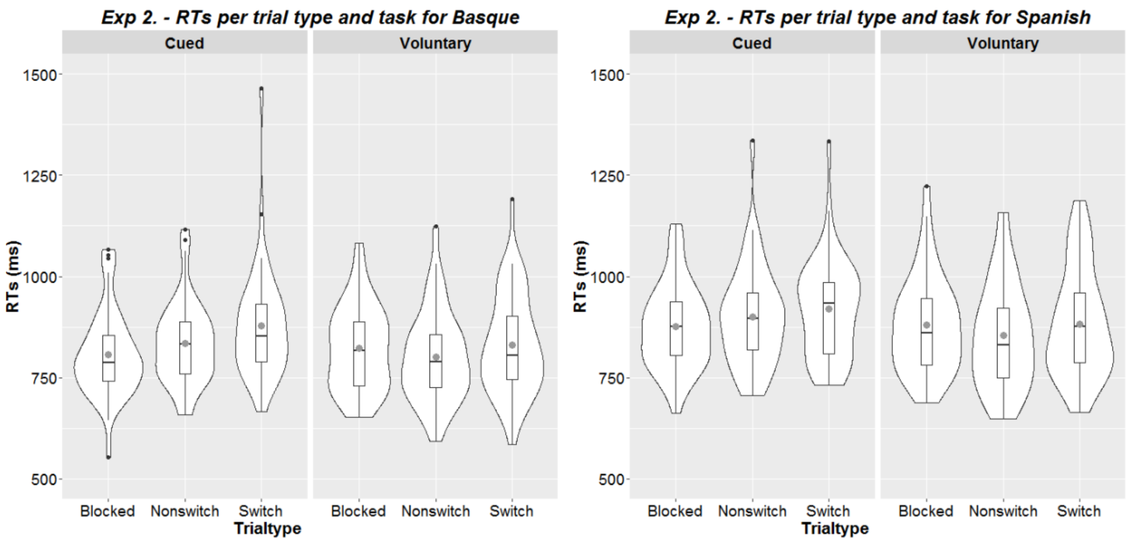

Figure 3. Violin plots showing the distribution of the RTs in Basque (left panel) and Spanish (right panel). For each language, the three trial types (blocked, non-switch, and switch trials) are shown for the cued (left) and voluntary (right) task. The boxplot shows the interquartile range, the thin black line the 95% confidence interval. The median is indicated by the thin black line while the grey dot represents the mean.

Discussion

Experiment 2 showed that voluntary and cued switching tasks yielded comparable switching costs. In contrast, mixing effects differed between the two task conditions, with a mixing benefit during voluntary switching and a mixing cost during cued switching. Similar to Experiment 1, there were no interactions with language. Furthermore, the voluntary mixed condition produced faster response times than the cued mixed condition. Contrary to previous studies (e.g., Kleinman & Gollan, 2016), no effects of task order were observed. However, in our study, the cued and voluntary tasks were completed on separate days, which may have minimised order effects.

There are several differences between cued and voluntary switching tasks that hinder a direct comparison of overall RTs and switching costs between the two tasks. As previously shown, switching costs in language and task switching can depend on various task-related features (cf. Declerck, Grainger, Koch, & Philipp, 2017). For example, we used different stimuli in the cued and voluntary tasks. However, stimuli were comparable between the two tasks as they were matched on frequency and length and showed similar RTs in the blocked conditions. Furthermore, using 30 stimuli per task minimised effects of individual stimuli and the models included item slopes for trial type to take into account that different items may respond differently to trial type. Both in the current study as well as in previous studies, the cued and voluntary tasks differ in other aspects too (e.g., cue monitoring and the percentage of switch trials).

Examining participant-related and item-related factors that could affect voluntary language mixing and switching

In addition to assessing the effects of voluntary versus cued tasks on mixing and switching costs, this study aimed to examine multiple potential factors (related to the bilinguals and/or the items) that could affect a bilingual's switching behaviour as well as the costs and benefits associated with language switching and mixing. Towards this end, we conducted additional analyses across the 96 participants who completed the voluntary switching task in Experiments 1 and 2. Two analyses focused on language choice and trial type as the dependent variables (DVs), to examine when and why bilinguals choose a specific naming language and switch between their languages. Two further analyses with RTs as the DV assessed several potential factors that could affect the switching and mixing effects.

Based on the literature, we examined three main types of possible predictors that could affect language choice and switching and/or the switching and mixing effects. First, focusing on language choice and the moment of switching, previous studies (e.g., Gollan & Ferreira, 2009) have suggested that there is a link between lexical access and language choice/switching. In order to examine this possible relationship more directly, the first two analyses on language choice and trial type included the RTs from the first Basque block as an indicator of the speed with which the items were accessed in Basque by each participant on the first presentation (i.e., the first time they named that item). This allowed us to examine whether lexical access can indeed predict language choice and switching. Second, balanced bilinguals have been found to switch more often than unbalanced bilinguals (Gollan & Ferreira, 2009). We therefore examined whether language use and proficiency predicted language switching behaviour as well as the switching and mixing effects. Lastly, inhibitory control has been argued to play an important role in cued language switching (e.g., Green, 1998) and correlations have been observed between cued switching costs and inhibition costs (e.g., de Bruin et al., 2014). To further examine whether voluntary switching is costly and related to inhibitory control and cognitive control more generally, measures of inhibitory control and working memory were assessed.

Analysis

To examine the role of these possible predictors and their effects on language choice and switching, we first constructed models with language choice and trial type (switch versus non-switch) as the dependent variables (Model 1 and 2). While language choice and trial type are related, they are not identical (e.g., a participant may have produced many non-switch trials in Basque only or in both Basque and Spanish). Furthermore, not all predictors are relevant for both DVs (e.g., inhibition costs could explain how often bilinguals switch but there is no clear theoretical motivation to argue that inhibition costs should predict whether an item is named in Basque or Spanish). The two models with language choice (Basque/Spanish) and trial type (switch/non-switch) as DVs only included mixed trials. As these are binary variables, they were fitted with generalized linear mixed-effect models (binomial family, 'bobyqa' optimiser). Next, we conducted two additional linear mixed-effect analyses, with RTs as the DV, to assess different factors possibly affecting the mixing and switching costs. More general details about these analyses can be found in the data analysis description for Experiment 1. All predictors (both significant and non-significant) are reported and predictors were not included or excluded based on the results.

For all predictors related to proficiency, use, and lexical access, we included the measurements related to Basque as this language showed more variance across participants (e.g., Spanish proficiency was at or close to ceiling for all participants). As a measurement of lexical access, we took the Basque RTs for each participant-item combination from the first Basque blocked condition. Only accurate responses were included (before outlier removal), which caused 5.4% of the data rows to have missing values[8]. As a measurement of Basque proficiency, we included the scores from the picture naming task as this task was completed by all participants and was most similar to the experimental task. As a measurement of language use, we included the self-rated percentage of Basque speaking time. Additionally, as a measurement of switching frequency, we included the self-rated percentage of switching within a sentence as this is most similar to the experimental switching task.

The following models were constructed, with all continuous predictors z-scored:

Model 1, DV: Language choice

Model 1 assessed effects of lexical access and Basque proficiency and use on language choice. It included main effects of trial type, blocked Basque RTs, Basque use and proficiency, and relative item length and frequency (the relative length and frequency of words in Basque compared to Spanish, following the prediction that words that are relatively longer or less frequent in Basque may be more likely to be

[8] For this reason, we use Basque blocked RTs rather than a relative measure of Basque versus Spanish RTs, which would have led to 9.5% missing values.

named in Spanish). In addition, we included block order (Spanish versus Basque first), as this could modulate effects of blocked Basque RTs. The model also included the following interactions: trial type x blocked Basque RTs; trial type x block order; blocked Basque RTs x block order; trial type x Basque proficiency and trial type x Basque use; trial type x item length and trial type x item frequency; trial type x blocked Basque RTs x block order. The full model did not converge, even after removal of random correlations. After removal of item slopes explaining the least variance, the final model (correlations removed) included all relevant subject slopes (trial type, blocked Basque RTs, item length, item frequency, and the corresponding interactions) and item slopes for Basque proficiency, block order, blocked Basque RTs x block order, and trial type x block order.

Model 2, DV: Trial type

The second model focused on trial type as the dependent variable and was similar to the model on language choice. However, we were particularly interested in assessing whether switching frequency was affected by inhibitory control and working memory. We therefore included the following main effects: language, blocked Basque RTs, block order, Basque proficiency and use, self-rated switching frequency, verbal Stroop cost and numerical Stroop cost as measurements of inhibitory control, and Digit span and Corsi span as measurements of working memory span. As interactions, we included language x blocked Basque RTs, language x Basque use, language x Basque proficiency, language x block order, blocked Basque RTs x block order, and language x blocked Basque RTs x block order (similar to the previous model). The final converging model (after removal of correlations) included intercepts with all relevant subject slopes (language, blocked Basque RTs, and their interaction) and item slopes for language, self-rated switching frequency, numerical Stroop costs, Digit span, block order, language x blocked Basque RTs, language x block order, blocked Basque RTs x block order, and language x Basque use.

Models 3 (mixing benefits) and 4 (switching costs), DV: RTs on the voluntary task

The third and fourth models examined effects of Basque proficiency and use, inhibitory control, and working memory on the mixing benefit and switching cost.

These models included the following main effects: trial type (mixing and switching costs); language; Basque proficiency; Basque use; self-rated switching frequency; verbal Stroop costs; numerical Stroop costs; Corsi span; Digit span. As interactions, we included trial type x language; Basque proficiency x trial type; Basque proficiency x language; Basque proficiency x trial type x language; Basque use x trial type; Basque use x language; Basque use x trial type x language; self-rated switching frequency x trial type; trial type x verbal Stroop costs; trial type x numerical Stroop costs; trial type x Corsi span; trial type x Digit span. For both models, the maximal random effects structure converged after removal of correlations (for subject, the only within-participant manipulations were trial type and language).

Results

Examining factors affecting language choice and language switching

Language choice

The first model focused on language choice as a DV (Basque coded as 0; Spanish coded as 1) to determine whether language choice was linked to lexical access (measured by the speed with which participants responded to each item the first time they named it in the Basque blocked condition). We included word frequency and length to examine whether items that were assumed to be easier to name in one of the languages (i.e., relatively shorter and higher frequency) were also named more often in that language. Lastly, we included Basque proficiency and Basque use to examine whether participants with a higher proficiency in and use of Basque also used Basque more often. The results of the full model are reported in Table 3.

Table 3. *Results of model 1, DV language choice (Basque scored as 0). For each predictor, the estimate, standard error, and z values are given. An asterisk indicates a significant effect. Non-switch trials were coded as -0.5; switch trials as 0.5. Word frequency and length are relative measurements, with higher values indicating relatively longer and more frequent Basque words.*

Predictor	**Estimate**	**SE**	**Z value**
Intercept	-0.286	0.060	-4.735*
Trial type	0.444	0.068	6.510*
Blocked Basque RT	0.187	0.027	6.864*

Basque proficiency	0.011	0.053	0.199
Basque use	-0.020	0.048	-0.418
Word frequency	0.067	0.052	1.307
Word length	0.029	0.054	0.541
Block order	-0.045	0.090	-0.504
Trial type x blocked Basque RT	0.086	0.027	3.135*
Trial type x Basque proficiency	0.105	0.084	1.252
Trial type x Basque use	0.015	0.083	0.182
Trial type x word frequency	0.015	0.026	0.573
Trial type x word length	0.016	0.027	0.611
Trial type x block order	0.187	0.143	1.310
Blocked Basque RT x Block order	-0.027	0.074	-0.361
Trial type x blocked Basque RT x block order	-0.049	0.055	-0.883

There was a main effect of trial type; in both Experiment 1 and Experiment 2, relatively more switches were made to Spanish than to Basque. In absolute numbers, a similar number of trials were switches to Spanish and Basque. However, there were more non-switch Basque trials than non-switch Spanish trials, leading to a relatively higher switching percentage in Spanish.

Lexical access affected language choice (see Figure 4, left panel showing a significant effect of Basque blocked RTs). The slower a participant named an item in Basque in the first block, the more often that participant named that item in Spanish in the voluntary condition. This relationship was only found for the individual RTs for each item-participant combination. When the average RT for each item across participants was used, there was no significant effect of Basque Blocked RTs on language choice (z = 0.147).

In addition to the main effect, an interaction was observed between Basque blocked RTs and trial type. Participants switched to Spanish more often when they named the item more slowly in Basque.

We examined whether a similar relationship between language choice and lexical access was also present for Spanish (Figure 4, right panel). The pattern is similar to what was observed for Basque: The slower the item was named in Spanish in the first block, the more often the item was named in Basque in the voluntary condition. This was also confirmed when Spanish blocked RTs instead of Basque blocked RTs were included in the model (main effect Spanish blocked RTs: z = -3.976). However, while there was a main effect of Spanish blocked RTs on language choice, this did not interact with trial type (z = 1.007). Thus, both Spanish and Basque RTs were related to overall language choice, with relatively more Basque responses for slower Spanish RTs and relatively more Spanish responses for slower Basque RTs. Slower Basque RTs were furthermore related to more switches to Spanish, but slower Spanish RTs led to more Basque responses in a similar manner for switch and non-switch trials.

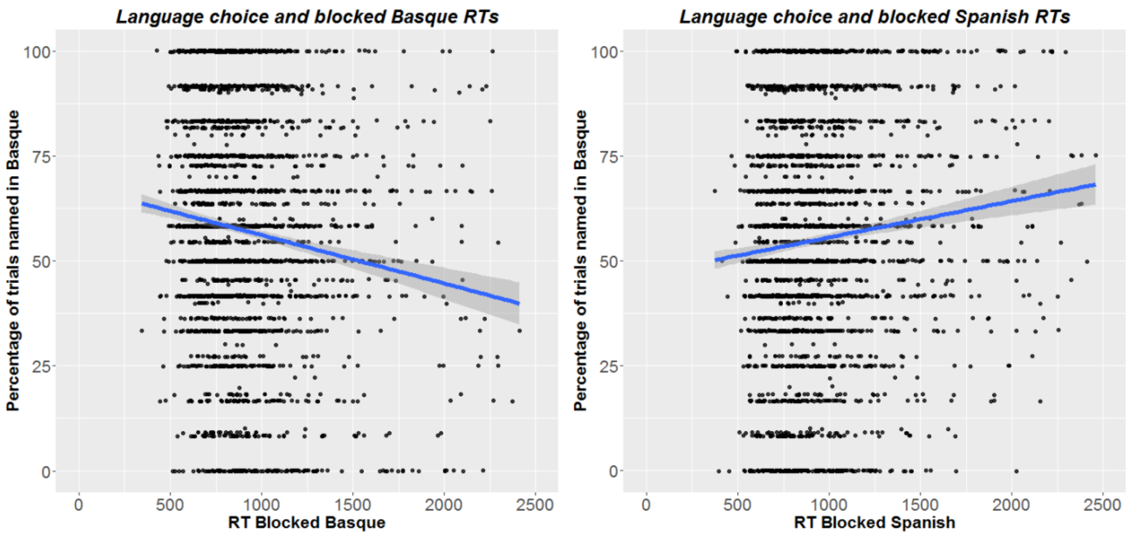

Figure 4. Percentage of trials named in Basque in relation to Basque blocked RTs (left) and Spanish blocked RTs (right). Each dot represents an individual item for an individual participant. To ensure comparability with the analyses, the percentage of trials named in Basque is based only on the items included in the analyses (i.e., correct responses only and excluding trials that could not be classified as switch or non-switch trials).

There were no main effects of Basque proficiency and use or any interaction with trial type, showing that participants with a higher proficiency in or use of Basque did not choose Basque more often. There were also no significant effects of word length and frequency, suggesting that language choice was not affected by general word characteristics.

The best model was the one including the significant predictors trial type, blocked Basque RT, and their interaction; removal of any effect resulted in a significantly worse model.

Trial type

The second model examined trial type as a DV (non-switch coded as 0; switch coded as 1; see Table 4 for the results from the full model). Similar to the previous model, we were interested in examining whether participants switched more often when lexical access was slower in the other language (as measured through blocked Basque RTs). We also included Basque proficiency and use, to assess whether more balanced bilinguals switched more often, as well as self-rated switching frequency in daily life. In addition, we examined whether switching frequency was related to inhibitory control and working memory span.

The basic findings of the model with trial type as a DV were similar to the model with language choice as a DV. There was a main effect of language, indicating that relatively more switches were made to Spanish than Basque. There was also a main effect of blocked Basque RTs, suggesting that participants switched more often when they were slower to name that item in Basque. Crucially, this interacted with language, such that relatively more switches were made to Spanish on items named more slowly in Basque. Similar to model 1, this interaction was not observed when Spanish blocked RTs were included (z = 0.682). Thus, while the participants' decision to switch to Spanish was related to how quickly they could name the item in Basque, slower Spanish naming speed was not related to more Basque switch than non-switch trials.

No main effects were observed of Basque proficiency or use, suggesting that balanced bilinguals did not switch more often than unbalanced bilinguals (if anything, they switched slightly less often). There was no significant relationship between self-reported language switching in daily life and switching frequency in the task. Furthermore, switching frequency was not predicted by the participant's inhibitory control scores or working memory span, suggesting that while inhibitory control may be involved in language switching, it is not directly related to the decision to switch voluntarily.

The simple model only including the significant predictors language, blocked Basque RT, and language x blocked Basque RT was the best model. Removal of any of these significant predictors led to a significantly worse model.

Table 4. *Results of model 2, DV trial type (non-switch scored as 0). For each predictor, the estimate, standard error, and z value are given. An asterisk indicates a significant effect. Basque trials were coded as -0.5; Spanish trials as 0.5.*

Predictor	**Estimate**	**SE**	**Z value**
Intercept	-0.296	0.043	-6.943*
Language	0.447	0.070	6.356*
Blocked Basque RT	0.031	0.014	2.301*
Basque proficiency	-0.087	0.052	-1.685
Basque use	0.021	0.052	0.399
Self-rated switching	0.057	0.044	1.304
Verbal Stroop costs	-0.075	0.045	-1.680
Numerical Stroop costs	0.024	0.044	0.552
Digit span	0.021	0.047	0.459
Corsi span	0.068	0.046	1.472
Block order	0.039	0.088	0.448
Language x Blocked Basque RT	0.088	0.027	3.265*
Language x Basque proficiency	0.126	0.086	1.469
Language x Basque use	0.002	0.086	0.019
Language x Block order	0.205	0.145	1.416
Blocked Basque RT x Block order	-0.019	0.028	-0.692
Language x Blocked Basque RT x Block order	-0.050	0.053	-0.948

Examining factors affecting the mixing benefit and switching cost

Experiments 1 and 2 both showed a switching *cost* but a mixing *benefit* in voluntary language switching tasks. In the current analysis, we examined whether these mixing benefits and switching costs were affected by the participants' Basque proficiency and use. In addition, we tested whether mixing/switching effects were related to performance on the inhibitory control and working memory tasks.

Tables 5 (mixing benefit) and 6 (switching cost) provide the full results of the statistical analyses. Consistent with the results from the individual experiments, the models on switching and mixing effects showed a main effect of blocked trials compared to non-switch trials that did not interact with language, reflecting the fact that the mixing benefit was similar for Basque (*M* = -41.1, *SD* = 88.2) and Spanish (*M* = -48.9, *SD* = 97.0)[9]. Switch trials were slower than non-switch trials, which did not interact with language, indicating a comparable switching cost for Basque (*M* = 35.4, *SD* = 48.0) and Spanish (*M* = 32.6, *SD* = 60.0).

In both models, Basque proficiency, Basque use, and self-rated switching frequency did not significantly predict overall RTs, mixing benefits, or switching costs.

With respect to the inhibitory control and working memory tasks, an interaction was found between switching costs and verbal Stroop costs. Participants with smaller verbal Stroop costs (i.e., assumed to reflect better inhibitory control) showed smaller language switching costs. This effect was not found for the non-verbal numerical Stroop cost. No further significant effects related to the background tasks were found, either in the mixing benefit or switching cost model[10].

The best model of mixing benefits included main effects of trial type (mixing) and language. For the model of switching costs, the best model included main effects of trial type (switching), language, and an interaction between switching costs and verbal Stroop costs. Removing any of the significant effects led to significantly worse models.

---

[9] This mixing benefit only shows that mixed *non-switch* trials were faster than blocked trials, but does not necessarily mean that using two languages is faster than using one language. To test the latter conclusion, we also compared the mixed condition as a whole to the blocked trials (using Model 3, with 'mixing' defined as the mixed condition (coded as 0.5) versus the blocked condition (coded as -0.5). The significant effect of condition (β = -0.033, *SE* = 0.010, *t* = -3.215) showed that using two languages was in fact faster than one and this did not interact with language (β = -0.003, *SE* = 0.008, *t* = -0.376).

[10] We also examined whether the inhibition costs were related to cued language switching by running Model 3 and 4 on the cued data from Experiment 2. This showed a main effect of verbal Stroop costs (Model 4: *t* = 3.193) that interacted significantly with the switching costs (*t* =2.006). Similar to the voluntary task, there were no effects of the nonverbal Stroop or working memory tasks.

Table 5. *Results of model 3 (mixing benefit), DV log RTs. For each predictor, the estimate, standard error, and t values are given. An asterisk indicates a significant effect.*

Predictor	**Estimate**	**SE**	**T value**
Intercept	6.705	0.015	451.851*
Mixing	-0.048	0.010	-4.898*
Language	0.060	0.011	5.613*
Basque proficiency	0.001	0.014	0.097
Basque use	-0.001	0.015	-0.087
Self-rated switching	-0.020	0.012	-1.690
Verbal Stroop cost	0.007	0.012	0.595
Numerical Stroop cost	0.012	0.012	1.012
Digit span	0.005	0.013	0.415
Corsi span	-0.010	0.013	-0.772
Mixing x language	-0.005	0.009	-0.581
Mixing x Basque proficiency	0.021	0.011	1.862
Language x Basque proficiency	0.010	0.010	0.934
Mixing x Basque use	0.003	0.011	0.299
Language x Basque use	-0.015	0.010	-1.454
Mixing x Self-rated switching	0.002	0.009	0.267
Mixing x Verbal Stroop cost	-0.005	0.009	-0.574
Mixing x Numerical Stroop cost	-0.002	0.009	-0.203
Mixing x Digit span	-0.013	0.010	-1.286
Mixing x Corsi span	-0.015	0.010	-1.548
Mixing x language x Basque proficiency	-0.004	0.010	-0.392
Mixing x language x Basque use	0.001	0.010	0.079

Table 6. *Results of model 4 (switching costs), DV log RTs. For each predictor, the estimate, standard error, and t values are given. An asterisk indicates a significant effect.*

Predictor	Estimate	SE	T value
Intercept	6.699	0.016	427.158*
Switching	0.038	0.004	9.421*
Language	0.055	0.009	6.380*
Basque proficiency	0.008	0.017	0.475
Basque use	0.003	0.017	0.184
Self-rated switching	-0.021	0.014	-1.463
Verbal Stroop cost	0.009	0.014	0.654
Numerical Stroop cost	0.015	0.014	1.040
Digit span	-0.001	0.015	-0.093
Corsi span	-0.020	0.015	-1.316
Switching x language	-0.002	0.007	-0.332
Switching x Basque proficiency	-0.008	0.005	-1.672
Language x Basque proficiency	0.003	0.008	0.337
Switching x Basque use	0.005	0.005	0.955
Language x Basque use	-0.010	0.008	-1.349
Switching x Self-rated switching	-0.002	0.004	-0.497
Switching x Verbal Stroop cost	0.013	0.004	3.037*
Switching x Numerical Stroop cost	0.006	0.004	1.509
Switching x Digit span	0.0003	0.005	0.067
Switching x Corsi span	-0.005	0.005	-1.019
Switching x language x Basque proficiency	-0.011	0.008	-1.349
Switching x language x Basque use	0.009	0.009	1.032

The robust mixing benefit in the voluntary switching situation may be the result of participants being able to avoid naming slower items in the mixed condition, an option that is not available in the blocked condition (cf. Gollan & Ferreira, 2009). We therefore tested whether mixing benefits were due to language preference and avoidance of naming certain items in one of the languages. Keeping in mind that the consistency of naming language may not only differ between participants or items, but also for individual item-participant pairs, we defined language consistency on each item for each bilingual, with scores ranging from 0.5 to 1. A score of 0.5 meant that the item was named 50% of the correct trials in language A and 50% in language B, while 1 meant that the item was always named in the same language. Here we present the results from the mixing benefit analysis; analyses focusing on the switching effect (as well as further analyses examining the mixing effect across time) can be found in the supplementary materials. The RT analysis with main effects of language, trial type, and naming language consistency showed a main effect of trial type (β = -0.045, *SE* = 0.009, *t* = -5.152), that interacted with language consistency (β = -0.024, *SE* = 0.004, *t* = -5.836). Comparing items with a low naming consistency (50%-74%) to those with a higher consistency (75%-100%) showed a larger benefit for the high consistency items (*M by item* = -101.9, *SD* = 33.6), but also a significant benefit for low consistency items (*M by item* = -17.0, *SD* = 18.7; β = -0.021, *SE* = 0.008, *t* = -2.522). Thus, while language preference explains part of the mixing benefit, it is not the only underlying mechanism.

General Discussion

The current study aimed to examine when and why bilinguals voluntarily switch between languages. Early and highly proficient Spanish-Basque bilinguals switched frequently (over 40% of trials). Language choice was predicted by the speed with which participants were able to name the pictures in each of the languages (i.e., pictures that were named more slowly in Basque were named more often in Spanish and vice versa). Switching costs were observed on both the voluntary and cued tasks. However, while cued language use resulted in mixing costs, the voluntary condition showed a mixing benefit. This suggests that using two languages voluntarily may be less costly than having to use only one language.

Switching frequency

The average switching rate in the current study was 41%, an average that is between the 35% observed for balanced bilinguals in Gollan and Ferreira (2009) and the 48% reported by Blanco-Elorrieta and Pylkkänen (2017). Laboratory studies on voluntary language switching are likely to elicit higher switching frequencies than those observed in daily life (e.g., Fricke & Kootstra, 2016, report codeswitching on 5.8% of the utterances in a corpus of Spanish-English bilinguals). The switching rate observed in laboratory studies may furthermore depend on many different task- and design-related variables (e.g., on whether participants complete a one-language only condition prior to the voluntary condition; Gollan & Ferreira, 2009). The exact instructions and use of a picture-word familiarisation phase may further affect switching frequency. For instance, our instructions encouraged participants to not use the same language throughout the whole task (which differs from Gollan & Ferreira, 2009), which may have increased switching rates.

We also studied several participant-related factors that could affect switching frequency. For example, Gollan and Ferreira (2009) reported a higher switching frequency for more balanced bilinguals compared to unbalanced bilinguals. In the current study, however, switching frequency was not related to the participants' proficiency in or use of Basque. We recruited participants with a range of Basque proficiency scores, encompassing quantile 15 to 100 from a large-scale study assessing proficiency scores in the tested population (de Bruin et al., 2017). However, while our range included balanced bilinguals as well as more Spanish-dominant bilinguals, all bilinguals had a high proficiency in both languages, and actually responded faster in Basque than Spanish. This relatively narrow range may have reduced any effect of proficiency.

Switching frequency was also not related to inhibitory control and working memory scores or to self-rated switching frequency in daily life. Self-ratings of language switching behaviour may be a poor estimate of actual switching behaviour. Jylkkä, Soveri, Laine, and Lehtonen (submitted) compared self-reported switching frequency to the Ecological Momentary Assessment (EMA), in which bilingual participants reported their language switching frequency every two hours during a two week span. While there were some correlations between self-rated switches and EMA measurements, overall the study suggested that self-reported language switching behaviour lacks convergent validity.

Language choice and lexical access

Language choice was predicted by lexical access. Items that were named more slowly in Basque were named more often in Spanish and vice versa. These findings are in line with previous studies suggesting that there is a link between lexical access and language selection. For instance, theoretically easier items (e.g., items with a higher frequency or shorter length) have been reported to be named relatively often in the non-dominant language (e.g., Gollan & Ferreira, 2009; Gross & Kaushanskaya, 2015). In our study, there was no link between the accessibility of the items in general (i.e., their length, frequency, and how quickly they were named across participants), suggesting that language choice was not related to the general properties of words. Rather, language choice was related to the accessibility of a specific word for a given participant. This suggests that, at least in bilinguals with a high proficiency in both languages, language choice may be driven by individual differences in lexical access.

Language switching, like language choice, was related to how quickly individuals could name the items, at least in Basque. Items with slow Basque responses were named more often in Spanish, and in particular there were relatively more switch trials to Spanish. Slow Spanish RTs were related to more frequent Basque responses too, but in a similar manner for non-switch and switch trials. Thus while Basque responses predicted when participants switched to Spanish, switches to Basque were not predicted by Spanish responses. This could suggest that other factors were driving participants to switch to or stay in Basque, for instance to use the default language (which, for most participants, was Basque). This is also compatible with our finding that even the fastest Spanish items did not show a Spanish preference (see Figure 4) and with the relatively high number of Basque non-switch trials. There appears to have been a general preference for Basque that increased when Spanish RTs were slow and was only reversed when access was slow in Basque.

Lexical access vs. top-down processes in voluntary language switching

Previous studies have suggested that voluntary language switching is not only governed by bottom-up processes related to lexical access but also by more top-down processes (e.g., Gollan & Ferreira, 2009). Similarly, corpus-based studies (e.g., Fricke & Kootstra, 2016) have found evidence for bottom-up processes (e.g., switching and language choice being affected by language and lexical priming) but have also suggested a role for top-down processing. If bilinguals follow a pure bottom-up approach (i.e., if they switch every time a word is more easily available in the other language), switching frequencies should be higher in daily life than e.g., the 5.8% observed by Fricke and Kootstra (2016). Thus, in daily life, a bottom-up approach is probably not always feasible and other factors could play a role in deciding which language to use (e.g., wanting to return to or stay in the default language).

Our study provides further evidence that both bottom-up and top-down processes are involved in voluntary language switching. The link with lexical access showed that bottom-up processes play a role in language choice. Yet, the switching and mixing effects suggest that top-down processes may play a role during blocked single-language picture naming as well as during language switching.

Language switching costs

In both experiments, a switching cost was observed in the voluntary language switching task. This contradicts some previous studies (e.g., Blanco-Elorrieta & Pylkkänen, 2017) but is compatible with others (e.g., Gollan & Ferreira, 2009; Kleinman & Gollan, 2016). Recent corpus-based analyses have also observed longer speech durations in syllables preceding a codeswitch compared to a non-switch, suggesting that switching between languages may be costly in natural speech (Fricke, Kroll, & Dussias, 2016).

A relationship was observed between verbal inhibition costs and voluntary switching costs. In cued language switching, reactive inhibition has been proposed as one of the underlying mechanisms: bilinguals may reactively suppress the non-target language in order to produce the correct response in the target language (e.g., Green, 1998). Some previous studies have reported correlations between inhibition tasks and language switching tasks, as well as the involvement of brain mechanisms related to inhibitory control during language switching (e.g., de Bruin et al., 2014; Jackson, Swainson, Cunnington, & Jackson, 2001). The relationship observed with verbal Stroop costs in the current study is in line with these studies, although it should be emphasised that in the current study this relationship was only found for the verbal but not for the non-verbal Stroop cost. While this should be interpreted with caution, the presence of switching costs in voluntary language switching, in combination with their relationship with inhibition costs, suggests that even in voluntary switching some form of reactive inhibitory control mechanism may be used.

Other mechanisms may be involved too. The switching cost may be related to switching between lexicons and activation of another target language being costly.

Furthermore, the switching cost may reflect the 'waiting time' if bilinguals first search for the word in the non-switch language and only switch when they cannot retrieve the word quickly enough. This could be similar to the idea that a non-word decision is reached in a lexical decision task after unsuccessfully trying to find a word in the mental lexicon for a certain amount of time (e.g., the dual-route cascaded model, Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Lastly, switching costs may not (only) stem from the lexical level, but also from sub-lexical phonological and articulatory levels. A switch between languages, while beneficial at the lexical level, may still come with a switching cost in terms of articulatory preparation, even though Spanish and Basque share a similar phonological structure.

Language mixing costs versus benefits

The cued (language) switching task in Experiment 2 showed a mixing cost for non-switch trials as compared to blocked trials, consistent with previous language- and task-switching studies (e.g., Christoffels, Firk, & Schiller, 2007). In contrast, the voluntary tasks showed a mixing benefit, as responses were slower on the blocked trials than on the mixed non-switch trials. Previous studies have reported voluntary mixing benefits too, but only for unbalanced bilinguals in the non-dominant language (e.g., Gollan & Ferreira, 2009). Our study expands on these findings by showing that, at least in a bilingual society, mixing benefits can also be observed in more balanced bilinguals, and in both languages. There are many differences between the current study and previous ones that make it difficult to evaluate under which circumstances mixing benefits may or may not occur. Mixing two languages rather than staying in one may be easier for the bilinguals tested in the current study as they are living in a bilingual society (i.e., the Basque Country) in which using two languages and switching between them is very common. Furthermore, the large phonological overlap between Basque and Spanish may make language mixing easier and faster than in other language pairs (e.g., Spanish and English). In addition, there are many other differences in the task design (e.g., the use of a familiarisation phase, the exact instructions used, the number of stimulus repetitions, etc.) that may affect the mixing effect.

The mixing benefit in the current study was partly explained by language preference, with larger mixing benefits for items with a more consistent naming language. However, the data also suggest that language choice and lexical access cannot fully account for the mixing benefit. First, we observed a mixing benefit for both languages, not just for the non-dominant language. Furthermore, while the mixing benefit was larger for items with a strong language preference, it was also observed for items without a strong preference. Thus, although the mixing benefit may be partly driven by bottom-up language choices based on lexical access, it suggests that using two languages is generally less costly than having to use only one language.

Indeed, several other studies not directly looking at language switching have suggested that using two languages is more beneficial than using one language. For instance, when bilinguals were allowed to use both languages, they retrieved more words in a tip-of-the-tongue study than when they could only use one language (Gollan & Silverberg, 2001). Similarly, performance on picture naming tasks such as the Boston Naming Test (BNT) has been found to be higher when bilinguals are allowed to use both languages instead of just one (e.g., Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007).

Mixing costs (and benefits) reflect the more global processes of maintaining and using two languages. In cued tasks, these costs are typically interpreted as the effort needed to maintain and use two tasks or languages (e.g., Rubin & Meiran, 2005). In these tasks, additional demands are placed on bilingual language control when two languages are needed compared to only one language. However, without the additional demands introduced by cued tasks, freely using two languages may be less effortful than having to stay in one language, as suggested by the mixing *benefit.* Even when only one language is used as the target language, the non-target language remains active and may cause interference (e.g., Thierry & Wu, 2007). As a consequence, the blocked condition in which only one language is used requires suppression of the non-target language. In contrast, in a mixed voluntary condition, both languages are allowed to be used freely and as such a lower global level of language suppression and control may be needed. While speculative, it could be argued that more proactive inhibition (i.e., more global inhibition of the non-target language in anticipation of only using one target language in the task) needs to be applied during the blocked than voluntary mixed condition. This interpretation would be in line with the Adaptive Control Hypothesis (Green & Abutalebi, 2013). Their single language context (which is most comparable to our blocked condition) is assumed to require increased control in terms of goal maintenance and interference control (encompassing conflict monitoring and interference suppression). In contrast, dense code-switching conditions (allowing free switching) do not require this additional control. The mixing benefits observed in the current study are compatible with the interpretation of increased global goal maintenance and inhibition during single-language blocks that is not needed in conditions that allow free switching.

Conclusion

While many studies have investigated cued language switching, relatively little work has been done on voluntary language switching, even though the latter would help to better understand language switching in more natural settings. The current study showed that proficient bilinguals frequently switched between their languages in a picture naming task, with language choice being related to lexical access. Furthermore, a mixing benefit was observed in the voluntary condition, suggesting that having to stay in one language may be more effortful than freely using two languages. This may be why bilingual language switching happens voluntarily in daily life.

Acknowledgement

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 743691). Support was provided by a Ministerio de Ciencia e Innovación Grant #PSI2014-53277, #PSI2015-65689-P, and by Ayuda Centro de Excelencia Severo Ochoa SEV-2015-0490. The authors would like to thank Itziar Basterra and Amets Esnal for their help with data collection. We would also like to thank Daniel Kleinman and two anonymous reviewers for their helpful comments.

References

Antón, E., Fernández García, Y., Carreiras, M., & Duñabeitia, J. A. (2016). Does bilingualism shape inhibitory control in the elderly? *Journal of Memory and Language*, *90*, 147-160.

Baayen, R. H. (2008). Analyzing Linguistic Data: A practical introduction to statistics using R. Cambridge, U.K.: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version*, *1*(7), 1-23.

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual Language Switching in the Laboratory versus in the Wild: The Spatiotemporal Dynamics of Adaptive Language Control. *Journal of Neuroscience*, *37*, 9022-9036.

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*, 192-208.

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

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*, 491-511.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*, 665-671.

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00522

de Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. *NeuroImage*, *90*, 348-359.

Declerck, M., Grainger, J., Koch, I., & Philipp, A. M. (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language*, *95*, 138-145.

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. *The Quarterly Journal of Experimental Psychology*, *71*, 808-816.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited. *Experimental Psychology*, *61,* 234-251.

Frank, A. F. (2011) R-hacks/mer-utils.R. https://github.com/aufrank/R-hacks/blob/master/mer-utils.R. Accessed October 2017.

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, *91*, 181-201.

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.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models* (Vol. 1). New York, NY, USA: Cambridge University Press.

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, *13*, 197-208.

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 640-665.

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*, 2167-2195.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*, 63-83.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515-530.

Gross, M., & Kaushanskaya, M. (2015). Voluntary language switching in English–

Spanish bilingual children. *Journal of Cognitive Psychology*, *27*, 992-1013.

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*, 169-178.

Jylkkä, J., Soveri, A., Laine, M., & Lehtonen, M. (submitted). Assessing bilingual language switching behavior with Ecological Momentary Assessment.

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*, 700-714.

Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 173,* 115-132.

Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, *15*, 651-662.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure?. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 575-599.

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*, 362-372.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, *40*, 25-40.

Milroy, L., & Muysken, P. (Eds.). (1995). *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. Cambridge, U.K.: Cambridge University Press.

Myers-Scotton, C., & Jake, J. L. (1995). Matching lemmas in a bilingual language competence and production model: Evidence from intrasentential code switching. *An Interdisciplinary Journal of the Language Sciences, 33*, 981-1024.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8-13.

Perea, M., Urkia, M., Davis, C. J., Agirre, A., Laseka, E., & Carreiras, M. (2006). E-Hitz: A word frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). *Behavior Research Methods*, *38*, 610-615.

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*, 859-862.

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods*, *49*, 1864-1881.

Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1477-1491.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*, 12530-12535.

Appendix A. Stimuli used in Experiment 1

Basque and Spanish words were matched on number of phonemes (Spanish: $M = 5.6$, $SD = 1.2$; Basque: $M = 5.6$, $SD = 1.6$; $t(29) = .297$, $p = .769$), number of syllables (Spanish: $M = 2.5$, $SD = 0.6$; Basque: $M = 2.5$, $SD = 0.7$; $t(29) = -.273$, $p = .787$), and log frequency (Spanish: $M = 1.3$, $SD = 0.5$; Basque: $M = 1.3$, $SD = 0.5$; $t(29) = -.259$, $p = .798$). Word length and frequency were determined through E-Hitz for Basque (Perea et al., 2006) and B-Pal for Spanish (Davis & Perea, 2005).

Spanish	**Basque**	**English**
Ardilla	Urtxintxa	Squirrel
Barba	Bizar	Beard
Boina	Txapel	Beret
Bombero	Suhiltzaile	Fireman
Burro	Asto	Donkey
Caballo	Zaldi	Horse
Cadena	Kate	Chain
Calcetín	Galtzerdi	Sock
Camisa	Alkandora	Shirt
Cangrejo	Karramarro	Crab
Cebolla	Tipula	Onion
Falda	Gona	Skirt
Flecha	Gezi	Arrow
Fresa	Marrubi	Strawberry
Luna	Ilargi	Moon
Mesa	Mahai	Table
Moneda	Txanpon	Coin
Montaña	Mendi	Mountain
Muñeca	Panpina	Doll
Nariz	Sudur	Nose
Pájaro	Txori	Bird
Puente	Zubi	Bridge
Pulmón	Birika	Lung
Red	Sare	Net
Regalo	Opari	Present
Rodilla	Belaun	Knee
Rueda	Gurpil	Wheel
Timbre	Txirrin	Bell
Vaca	Behi	Cow
Vestido	Soineko	Dress

Appendix B. Stimuli used for the cued language switching task in Experiment 2

Basque and Spanish words were matched on number of phonemes (Spanish: $M = 5.7$, $SD = 1.7$; Basque: $M = 5.5$, $SD = 2.0$; $t(29) = .552$, $p = .585$), number of syllables (Spanish: $M = 2.5$, $SD = 0.7$; Basque: $M = 2.5$, $SD = 0.9$; $t(29) < .001$, $p > .999$), and log frequency (Spanish: $M = 1.2$, $SD = 0.4$; Basque: $M = 1.2$, $SD = 0.5$; $t(29) = .161$, $p = .873$). The voluntary and cued stimuli did not differ significantly in Basque number of phonemes ($t(29) = .143$, $p = .887$), number of syllables ($t(29) = -.162$, $p = .873$), or log frequency ($t(29) = .746$, $p = .461$). Similarly, there were no significant differences between voluntary and cued stimuli for Spanish number of phonemes ($t(29) = -.283$, $p = .779$), number of syllables ($t(29) = -.465$, $p = .645$), or log frequency ($t(29) = .545$, $p = .590$). Word length and frequency were determined through E-Hitz for Basque (Perea et al., 2006) and B-Pal for Spanish (Davis & Perea, 2005).

Spanish	Basque	English
Anillo	Eraztun	Ring
Bruja	Sorgin	Witch
Calvo	Burusoil	Bold (person)
Ceja	Bekain	Eyebrow
Cocina	Sukalde	Kitchen
Conejo	Untxi	Rabbit
Corazón	Bihotz	Heart
Corredor	Korrikalari	Runner
Cuerda	Soka	Rope
Desayuno	Gosari	Breakfast
Escoba	Erratz	Broom
Fuego	Su	Fire
Gafas	Betaurrekoak	Glasses
Guante	Eskularru	Glove
Horno	Labe	Oven
Jardinero	Lorezain	Gardener
Ladrón	Lapur	Thief
Llave	Giltza	Key
Manzana	Sagar	Apple
Molino	Errota	Mill
Murciélago	Saguzar	Bat
Oreja	Belarri	Ear
Pan	Ogi	Bread
Pato	Ahate	Duck
Queso	Gazta	Cheese
Rana	Igel	Frog
Serpiente	Suge	Snake

Toro	Zezen	Bull
Trigo	Gari	Wheat
Ventana	Leiho	Window