Bilingualism: Language and Cognition

cambridge.org/bil

Does language switching behavior rely on general executive functions?

Jussi Jylkkä[1,2], Matti Laine[1] and Minna Lehtonen[2,3]

[1]Department of Psychology, Åbo Akademi University, Finland; [2]Department of Psychology and Speech-Language Pathology, University of Turku, Finland and [3]Center for Multilingualism in Society across the Lifespan, Department of Linguistics and Scandinavian Studies, University of Oslo, Norway

Research Article

Cite this article: Jylkkä J, Laine M, Lehtonen M (2021). Does language switching behavior rely on general executive functions? *Bilingualism: Language and Cognition* **24**, 583–595. https://doi.org/10.1017/S1366728920000619

Received: 23 August 2019
Revised: 11 August 2020
Accepted: 21 September 2020
First published online: 5 November 2020

Keywords:
language switching; cued naming; Ecological Momentary Assessment; executive functions; task-specificity; task-generality; skill-learning

Address for correspondence:
Jussi Jylkkä,
E-mail: jjylkka@abo.fi

Abstract

The assumption that everyday language switching trains bilinguals' executive functions (EF) presupposes that language switching engages domain-general EF. This study examined associations between three types of tasks in a group of Finnish-English late bilinguals: everyday language switching frequency assessed with Ecological Momentary Assessment, language switching performance on a cued bilingual naming task, and EF as measured with Simon, Flanker, and Number-letter tasks. Cued switching performance showed mainly positive associations with EF performance, but the associations between everyday language switching and cued switching performance, and between everyday switching and the EF tasks were largely against our hypotheses. The findings indicate that participants with lower monitoring capacity make more everyday language switches. This speaks against the idea that everyday language switching would facilitate executive functioning. The results suggest that associations between language switching and general EF are more complex than current models assume.

1. Introduction

It has been suggested that bilingual experience could train general executive functions (EF) (e.g., Bialystok, 2009). More specifically, the bilingual training of EF is assumed to take place through certain language behaviors or processes utilizing general EF processes: for instance, inhibition of the non-target language, switching between two languages, and monitoring of the activation levels of the two languages are supposed to rely on task-general inhibition, shifting, and monitoring, respectively (e.g., Linck, Schwieter & Sunderman, 2012; Rodriguez-Fornells, de Diego Balaguer & Münte, 2006). We will call this assumption THE TASK-GENERALITY HYPOTHESIS (Jylkkä, 2017). However, several recent studies (Dick, Garcia, Pruden, Thompson, Hawes, Sutherland, Riedel, Laird & Gonzalez, 2019; Jylkkä, Soveri, Wahlström, Lehtonen, Rodríguez-Fornells & Laine, 2017; Paap, Johnson & Sawi, 2015; Paap et al., 2017; Paap & Sawi, 2014) and an extensive meta-analysis (Lehtonen et al., 2018) indicate that no such bilingual advantage exists. One possible reason for the lack of a bilingual executive advantage is that bilingual language behavior simply does not engage general EF, and therefore cannot train them. We examined this issue in the present study.

In contrast to the task-generality hypothesis, bilingual behaviors may rely on task-specific mechanisms, and only indirectly or not at all on task-general mechanisms. If this were the case, it would not be possible to train general EF through bilingual behavior. The "curse of task-specificity" is widely reckoned in cognitive training literature. For instance, training working memory (WM) with specific tasks, such as n-back, rarely if ever leads to enhancement in tasks dissimilar to the trained task: there is generally lack of transfer beyond the trained WM task paradigm (for a meta-analysis, see Soveri, Antfolk, Karlsson, Salo & Laine, 2017). This can be taken to suggest that learning to perform a specific WM task represents a variety of SKILL LEARNING, which is task-specific (Gathercole, Dunning, Holmes & Norris, 2019; Taatgen, 2013). As suggested by Chein and Schneider (2012), general executive resources are utilized only at the outset of novel task learning, after which enhancement in the task is mainly due to automatization in task routines and strategies, which are typically task-specific. Based on this skill-learning framework, Paap (2018) puts forward a "controlled dose hypothesis" regarding the relationship between general EF and bilingual behaviors: he proposes that general executive functions are utilized only when learning a novel language, and it is only during this phase that bilingual behavior could even in principle train general EF. On Paap's account, bilingual behavior moves towards automatic (not executively demanding) processing once learning advances. The lack of consistent bilingual advantage can be considered as indirectly supporting the task-specificity view. Task-generality in bilingualism has been more directly tested in experiments that correlate

© The Author(s), 2020. Published by Cambridge University Press

bilingual task performances against general (often non-verbal) executive task performances, which we will discuss later.

It can even be questioned whether specific executive tasks rely on any task-general executive mechanisms. This is known as the task-impurity problem in executive functions (Miyake et al., 2000): different EF tasks even in the same domain typically show low intercorrelations (e.g., Paap et al., 2015; Paap & Sawi, 2014), indicating that they do not (at least strongly) rely on the same underlying mechanisms. For instance, Shilling, Chetwynd, and Rabbitt (2002) report that different versions of the Stroop task, varying only with respect to the stimuli that were used, show non-existent or very small correlations (r's between -.13 and .22, N's between 33 and 49). Similarly, Salthouse (2010) found that two versions of the Flanker task, one with arrows and the other one with letters, did not correlate at all ($r = .03$, $N = 265$). Additionally, in Miyake et al. (2000), a well-cited study on the latent factor structure of commonly used executive tasks, residuals in specific tasks ranged from .78 to .94, indicating that most of the variance in the tasks was not explained by the latent factor. These findings can be taken to support a skill learning approach to executive tasks: learning to perform an executive task is always (at least to some extent) a task-specific skill, which only indirectly relies on common resources that can be utilized across different tasks. The common resource might simply be a general ability to learn new tasks (Jylkkä, 2017).

Here we aimed to directly address the task-generality hypothesis in the case of language switching by examining associations between language switching behavior and performance on non-verbal executive tasks in healthy bilingual adults. We examined associations between three different domains: 1) everyday language switching frequency; 2) performance on a bilingual cued picture naming task; and 3) performance on general EF tasks. We hypothesized that performance in the cued naming task and in the EF tasks should be associated if they rely on the same cognitive mechanisms. The relationship between everyday language switching frequency and the laboratory tasks is more complex. Higher switching frequency does not necessarily reflect better EF or switching performance on a laboratory task, but we hypothesized that participants with better EF or switching performance in the lab would more frequently make intended language switches, and less often unintended switches (Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman & Münte, 2012). Another possibility is that everyday language switching would train both EF performance and language switching performance that can be measured in lab context, under the assumption that they rely on the same cognitive mechanisms. However, due to the correlational nature of the present study, these two hypotheses cannot be distinguished. Thus, we may endorse the conservative interpretation that if there were an association between everyday language switching frequency and performance in the laboratory tasks, this would be due to the laboratory tasks engaging the same mechanisms as everyday language switching. In other words, on this interpretation, a positive association between everyday language switching and the laboratory tasks would not necessarily reflect a training effect, but instead simply engagement: subjects with better EF or language switching performance switch more in everyday life because they can do so efficiently.

In sum, if all the three types of behaviors (performance in the EF tasks and in the cued naming task, and everyday language switching frequency) rely on the same task-general cognitive functions, we should find associations between them. To our knowledge, this is the first study to directly assess the task-generality hypothesis across the three domains.

1.1. Everyday language switching and EF

A central question in the case of language switching is which types of everyday language switching behaviors would engage general executive functions to the highest extent. Previous literature is inconsistent concerning the associations between everyday language switching and general EF task performance. In some studies, higher frequency of everyday language switching has predicted better executive performance (Hartanto & Yang, 2016; Prior & Gollan, 2011; Soveri, Rodriguez-Fornells & Laine, 2011; Verreyt, Woumans, Vandelanotte, Szmalec & Duyck, 2016), while others have failed to find such associations (Johnson, Sawi & Paap, 2015; Jylkkä et al., 2017; Paap et al., 2017; Yim & Bialystok, 2012). These study designs have been plagued by problems of ecological validity, because they have typically measured everyday language switching with questionnaires (e.g., the Bilingual Switching Questionnaire or BSWQ; Jylkkä et al., 2017; Soveri et al., 2011) or with simple individual questions (e.g., "How often are you in a situation in which you switch between languages?" (Verreyt et al., 2016; see also Johnson et al., 2015; Prior & Gollan, 2011)). In an earlier experiment (Jylkkä, Soveri, Laine & Lehtonen, 2019), we addressed this problem by utilizing Ecological Momentary Assessment (EMA), where the language switches made by a speaker were probed multiple times a day using a smartphone application. The results indicated that the retrospective language switching questionnaires and individual questions are not valid. However, we found evidence that contextual language switches (operationalized with the question "How large proportion of the aforementioned language switches took place in situations in which you typically always switch between languages?"), assessed with EMA, were associated with better executive functions.

Why would the frequency of everyday contextual language switching be associated with EF performance? One possible explanation is provided by the Adaptive Control (AC) hypothesis (Green & Abutalebi, 2013) which distinguishes between three interactional contexts: Single language context, Dense code-switching, and Dual language context. In a Single language context, a multilingual person utilizes only one language, and hypothetically this loads very little on their executive system: the speaker only needs to inhibit intrusions from the non-target language. In a Dense code-switching context, the pressures on the executive system are even smaller, because in that situation "anything goes": a speaker may use any language that is most readily available and utilize a control process that Green and Abutalebi call "opportunistic planning". In turn, in a Dual language context the executive system is expected to be utilized to the highest degree, because there the speaker uses one language with a specific person (or group of people), and another language with another person (or group of people). In this context, the speaker has to continuously switch between languages under the restriction that the speaker must use a specific language with a specific person. Hartanto and Yang (2016) found evidence for an association between higher frequency of language switching in dual language context (as measured with estimates calculated from retrospective self-reports) and smaller switch costs in a color-shape shifting task.

It could be argued that the finding in our previous EMA study (Jylkkä et al., 2019) where contextual switches were associated with better executive performance was because the EMA question concerning contextual switches tapped on switching in a Dual

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

language context. In our contextual switches question "How large proportion of the aforementioned language switches took place in situations in which you typically always switch between languages?", the participants may have interpreted the word "situations" as meaning persons with whom they typically use a specific language. In the present study, we examined this possibility by introducing the EMA application questions that tap on Dual language switching versus Dense code switching. Dual language switching was assessed by Between-Interlocutor Switches (BIS) that tapped on how many times the participant switched between different speakers, using one language with one person and another language with another person. Dense code switching was measured by Within-Interlocutor Switches (WIS) that tapped on how many times the speaker switched between languages while speaking to one person. Along the lines of the Adaptive Control hypothesis, we hypothesized that BIS would be more strongly associated with general executive performance than WIS.

1.2. Cued language switching performance in a laboratory task and general EF

If language switching relies on general executive functions, we would expect participants' performance in a laboratory-based bilingual cued picture naming task to correlate with their performance on general executive tasks. Moreover, it can be hypothesized that such associations would be stronger compared to associations between everyday language switching and computerized EF tasks, because both cued naming and the EF tasks are performed in the same context (laboratory) and are structurally similar. Previous studies focusing on the associations between laboratory-based language switching and general EF have, however, yielded inconsistent results. Linck et al. (2012) reported that better inhibitory capacity, measured with a Simon task, was associated with reduced language switch costs in the dominant language L1, in line with the task-generality hypothesis (and Green's (1998) inhibitory control model of language switching, in particular). However, a previous study by our group, utilizing multiple executive tasks, could not replicate this result (Jylkkä, Lehtonen, Lindholm, Kuusakoski & Laine, 2018b). We did find statistically significant associations between language switching and EF performance, but the results were inconsistent and not in line with any theoretical account. In the present study, we investigated whether any of the significant effects in the previous study would be replicable.

Discrepancies between bilingual switching and general executive tasks have been reported by Calabria, Hernández, Branzi, and Costa (2012), Calabria, Branzi, Marne, Hernández, and Costa (2015), Segal, Stasenko, and Gollan (2019), as well as Stasenko, Matt, and Gollan (2017). On the other hand, Declerck, Grainger, Koch, and Philipp (2017) found that performance in closely matched bilingual and general switching tasks were associated. Likewise, Struys, Woumans, Nour, Kepinska, and Van Den Noort (2019) found weak evidence for an association between language switching and general EF, but this was supported only by one positive finding out of 25 comparisons, which could be due to chance. Lastly, Timmer, Calabria, and Costa (2019) found that training on a bilingual switching task led to improvement in a general switching task. This could be taken to support the task-generality hypothesis, but it is also in line with the skill learning theories (and Paap's controlled dose hypothesis, in particular), which imply that general EF are employed only during initial stages of language learning.

1.3. Cued language switching performance and everyday language switching

In addition to the abovementioned two topics that both focus on associations between language switching and EF, in this study we also investigated associations between frequency of everyday language switching as assessed with EMA, and the participants' performance on a cued switching task in laboratory. We hypothesized that participants with better language switching performance, as measured with the cued naming task, would have higher frequency of demanding language switches: namely, intended switches (in contrast to unintended switches) and between-interlocutor switches (in contrast to within-interlocutor switches) (see Rodriguez-Fornells et al., 2012 and Green & Abutalebi, 2013, respectively). This was expected on the assumption that more efficient switchers would also switch more often in everyday life.

1.4. Summary of the setup and hypotheses

Our goal was to examine the task-generality hypothesis which implies that different tasks that hypothetically elicit executive load are correlated. To this end, we used three different types of measures: 1) everyday language switching assessed with EMA, 2) performance on a cued bilingual naming task in laboratory, and 3) performance on general executive tasks. The task-generality hypothesis implies that insofar as all these behaviors engage the same, general executive system, they should be correlated. We expected that higher frequency of everyday language switching would be associated with better performance in the cued naming task, because these two behaviors are of the same type (language switching) although they occur in different contexts (lab vs. everyday life). We also expected that cued naming would be associated with general executive performance, because both tasks are structurally similar computerized tasks. On the other hand, we hypothesized that the associations between everyday language switching and performance on the executive tasks would be weaker than the two aforementioned associations, because they take place in different contexts (everyday life vs. lab) and represent different domains (verbal vs. non-verbal).

In turn, the task-specificity hypothesis can be formulated along the lines of the skill-learning approach of Chein and Schneider (2012). It predicts that the more experience one has in a task, the more that task is governed by automatized, task-specific mechanisms. Thus, the task-specificity approach predicts that associations would be stronger between the lab-based cued naming and EF tasks, because both include novel features for the participant and therefore rely on overlapping task-general mechanisms. In contrast, everyday language switching would on this reasoning be an automatized process that largely relies on task-specific mechanisms and would only weakly correlate with the cued switching and EF tasks. The setup and hypotheses are summarized in Figure 1.

2. Method

2.1. Participants

The participants (N = 33, 28 female) were late Finnish-English bilinguals who had learnt L2 (English) after the age of 7 at school (mean self-rated age of acquisition 7.55, SD = 2.29). Their self-reported overall proficiency in L2 was on average 6.24 (SD = .66) on a scale from 1 to 7. Most of the participants had either a

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

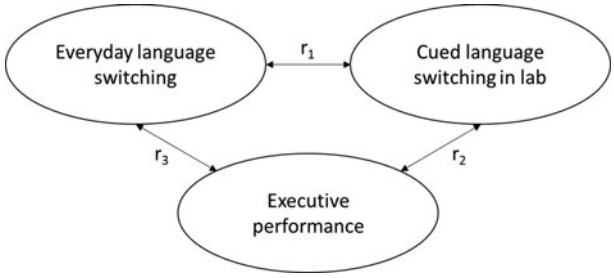


Fig. 1. The task-generality hypothesis implies behavior in all the three domains is associated, because they rely on task-general executive functions. Moreover, we expected that r1 and r2 would be higher than r3, because everyday language switching and cued naming are both in the verbal domain although they take place in different contexts, and cued naming and EF performance are in the same context albeit in different domains (verbal vs. non-verbal), whereas everyday switching and executive performance take place both in different domains (verbal vs. non-verbal) and in different contexts (lab vs. everyday life). The task-specificity or skill learning hypothesis, in turn, implies that the novel laboratory tasks (cued switching and EF-tasks) would be more highly correlated with each other than everyday language switching would be with either, because everyday language switching is an automatized process that relies on task-specific cognitive mechanisms.

bachelor's degree (n = 12) or a master's degree (n = 19), while one subject had a degree from a vocational high school and one had a doctoral degree. Eight of the participants (24%) reported using more languages than just Finnish and English: namely, Swedish (n = 6), German (n = 3), Russian (n = 2), Spanish (n = 1), Dutch (n = 1), Czech (n = 1), and French (n = 1) (4 participants reported 1 extra language, 2 reported 2, 1 reported 3, and 1 reported 4). Their average age at the time of testing was 26.70 years (SD = 4.59). All were native Finnish speakers, and neurologically healthy. They were recruited through e-mail lists and social media primarily from the University of Turku and the University of Helsinki (both of which are dominantly Finnish-language universities) and received movie tickets as compensation.

2.2. Procedure

The study was approved by the Joint Ethics Review Board of the Departments of Psychology and Logopedics at the Åbo Akademi University. After enrolling for the study, the participants came to the laboratory where they received information about the study and performed the executive tasks (Simon, Flanker, and Number-Letter) and the bilingual cued naming task. Task presentation was counterbalanced across participants so that half of the participants received the EF tasks first and the other half received the cued naming task first. Additionally, the order of the individual EF tasks, as well as the order of the mixed and single blocks in the cued naming task, was counterbalanced. After the EF and cued naming tasks, the participants received written instructions about how to use the EMA application, including information about the intended meaning of the questions with practical examples. After that, they installed the EMA application on their smartphone and started the EMA period where they would assess their language switching behavior on a daily basis.

Executive tasks

The Simon task was used as a measure of inhibition or conflict resolution (Simon & Rudell, 1967). The participant had to categorize the color of the stimulus (blue or red) by pressing either a left or right button, respectively. The stimuli appeared randomly on the left or right side of the screen. On congruent trials, the stimulus was on the same side as the response button (e.g., red stimulus on the right), whereas on incongruent trials the stimulus appeared on the side opposite to the correct response button (e.g., red stimulus on the left). Hypothetically, on incongruent trials the participant needs to inhibit the irrelevant information about the spatial location of the stimulus. The present version of the test consisted of 100 trials, half congruent and half incongruent. Each trial began with a fixation cross (800 ms) followed by a 250 ms blank interval. After that, the stimulus appeared and remained on the screen for 1,000 ms unless a response was given. Finally, the screen was blank for 500 ms. The stimuli were presented in four blocks with 5-second intervals in-between. The Simon effect is the difference between congruent and incongruent trials, with a larger effect reflecting worse performance. In the analyses performed with trial data, congruency effect was defined as the contrast between the congruent and incongruent condition.

The Flanker task (adapted from Eriksen & Eriksen, 1974) can be considered as a measure of inhibition or selective attention. In this task, the participant was presented with an array of five arrows, and their task was to categorize the direction of the central arrow by pressing a button. On congruent trials, all arrows pointed in the same direction (e.g., >>>>>), whereas on incongruent trials the central arrow pointed in the opposite direction from the "flankers" (e.g., >><>>). The flanker effect is the difference between congruent and incongruent trials, with a larger difference reflecting worse performance. In the present variant of the task, there were 100 trials, half of which were congruent and half incongruent. Each trial began with a fixation cross (800 ms), followed by the stimulus that remained on the screen for 800 ms unless a response was given, and finally a blank screen (500 ms). The stimuli were presented in four blocks with 5-second intervals in-between.

The number-letter task (adapted from Rogers & Monsell, 1995) was assumed to tap on non-verbal task switching (or cognitive flexibility). In this task, the participant was presented with a number-letter pair (e.g., A3) in one of two vertically aligned boxes. The participant was instructed to categorize the number as odd (response button one) or even (response button two) if the pair appeared in the upper box, and the letter as vowel (response button one) or consonant (response button two) if the pair appeared in the lower box. The present version of the task consisted of two single task blocks that included 32 trials each (letter categorization only or number categorization only), and one mixed block (80 trials). In the mixed block, there were 32 switch trials and 47 repetition trials (the first trial is neither). On switch trials, the task (location of the stimulus) switched from the previous trial, whereas on repetition trials, the task (location) stayed the same. Each trial began with a blank interval (150 ms), followed by a fixation cross (300 ms). After that, the stimulus appeared and remained on the screen for 3,000 ms unless a response was given. There was a short break between each block. The task yields two measures: a switch cost, which is the difference in RT between switch and repetition trials, and a mixing cost, which is the difference between repetition trials and single block trials, generally assumed to tap on monitoring. In both measures, a larger difference indicates worse performance.

In all EF tasks, only reaction times from correct trials were used in the analyses.

EMA method

We employed five questions about language switching in speech: Intended Switches (EMA-IS), Unintended Switches (EMA-US), the proportion of L1 speech during the last two hours (EMA-Balanced), Within-Interlocutor switches (EMA-WIS),

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

Table 1. Language switching questions used in the EMA method.

Abbreviation	Question	Response scale	Re-coding for analysis	Average† (SD)	Min; Max
EMA-IS	During the last two hours, how many times did you intentionally switch between languages in your speech?	0, 1–3, 4–6, 7 or more times	0, 1, 2, and 3, respectively	.64 (.40)	.038; 1.57
EMA-US	During the last two hours, how many times did you unintentionally switch between languages in your speech?	0, 1–3, 4–6, 7 or more times	0, 1, 2, and 3, respectively	.33 (.28)	0.00; .92
EMA-Balanced*	How large a proportion of your speech during the last two hours was in Finnish?	Continuous from 0–100%	No re-coding	.31 (.21)	.010; .72
EMA-WIS	During the last two hours, how many times did you switch between languages while speaking with the same person?	0, 1–3, 4–6, 7 or more times	0, 1, 2, and 3, respectively	.56 (.35)	.027; 1.55
EMA-BIS	During the last two hours, how many times did you switch between languages between two different persons? (First speaking in one language to one person, then in another language to another person.)	0, 1–3, 4–6, 7 or more times	0, 1, 2, and 3, respectively	.44 (.38)	.038; 1.82

* The EMA-balanced question was not included in the analyses.† Average as well as minimum and maximum values were calculated from individual means. Individual observations from EMA were not used, because then the overall averages would have been affected by differences in response rates between participants.

and Between-Interlocutor Switches (EMA-BIS; see all questions and their codings in Table 1). The first two questions were adapted from our previous study (Jylkkä et al., 2019), whereas the last three were designed to assess the AC hypothesis. We assumed that EMA-WIS would tap mainly dense code switching, whereas the EMA-BIS would tap switching in a dual language context. The EMA-Balanced question was included to assess the extent to which the participants used only one language versus both languages during the last two hours. This question was, however, omitted from the analyses because of a lack of theoretical motivation, and to reduce the number of analyses and thereby minimize the risk of false positives.

The EMA period lasted for 14 days, and the participants received automatic notifications to answer the questions 6 times a day between 9 am and 9 pm. As the questions were about language switching during the last two hours, they covered a whole 12 hour period each day. The participants gave on average 3.39 responses per day (SD = .69), and the average length of the EMA period was 12.97 days (SD = 1.80). Average responses are summarized in Table 1.

The cued naming task

The picture naming task was identical to the one used in Jylkkä et al. (2018b). Participants named aloud photographs of ordinary objects in Finnish or English, depending on a color cue. The responses were recorded for analysis. Each object was isolated on a white rectangular background and standardized in size. The cue color formed the background of the whole screen, on which the photo on the white background was centered. The participant was to name the picture in English if the background was red, and in Finnish if the background was blue. The cues were pseudorandomized so that there were always 2-4 consecutive same-language trials. A trial began with a blank white screen (1,000 ms), followed by a black fixation cross in the center of the white screen (500 ms). Then the stimulus picture on the cue-colored background was shown for 1,500 ms (irrespective of whether a response was given). The reaction times were determined through computerized analysis of audio recordings of the responses, based on amplitude (for details, see Jylkkä et al., 2018b). The correctness of responses was registered manually by a research assistant, and only correct trial RTs were used in analyses.

To obtain a mixing cost measure, two single language blocks (English and Finnish) were used in addition to a mixed language block. The single language blocks included 90 trials each and the mixed language block 180 trials. We had altogether 90 photos as stimuli, and their names in Finnish and English were matched in terms of length in phonemes and lemma frequency (p's > .5). The single language blocks consisted of 90 trials, i.e., all the photos in random order. The same photos were used in both single language blocks. The mixed language block consisted of 180 trials, i.e., the 90 photos named once both in Finnish and in English. Every photo was thus presented twice in the mixed block. There was at least a 10-trial interval before a photo was presented for the second time. The naming language was assigned randomly to each photo, i.e., the stimuli were bivalent. The mixed language blocks included 119 repetition trials and 60 switch trials (in sum 179 switch and repetition trials; the first trial is neither a repetition nor a switch trial), with an equal number of switches into both languages. We had 4 variants of each block, each separately pseudorandomized. The presentation order of the block variants was counterbalanced between subjects.

The cued naming task yields two measures. The language switch cost, i.e., difference between repetition and switch trials within the mixed block, is typically assumed to reflect (executive) control processes in lexical access (Green, 1998; Meuter & Allport, 1999). The language mixing cost, i.e., difference between repetition trials in the mixed block and single block trials, can be interpreted as a monitoring or preparedness cost: even though the language stays constant during the mixed block repetition trials, the participant needs to be prepared for a possible switch, unlike in the single language block. More specifically, the mixing cost may be related to maintaining two competing task schemas in working memory, which requires executive monitoring as to which schema to use (Botvinick, Braver, Barch, Carter & Cohen, 2001; Koch, Prinz & Allport, 2005; Prior & MacWhinney, 2010; Rubin & Meiran, 2005).

3. Results

All analyses were run in R using multilevel models (package lme4). Because it is only possible to have the dependent variable in long format in a multilevel model (i.e., including each trial in the analysis), the independent variables in each model were

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

inserted as means. The use of multilevel models was motivated by the small size of the sample: multilevel models increase statistical power through including each trial as an observation (in contrast to having one overall mean per participant).

3.1. Everyday language switching and cued picture naming

Measures of everyday language switching frequency are henceforth titled ES-[type] (type = IS, US, WIS, and BIS), and the cued naming measures are titled CN-[type] (type = switching cost and mixing cost).

In the CN task, all trials deviating more than +/−3SD from the participant's mean were excluded as outliers; this resulted in the removal of 116 trials (1.11% of total trials). In the ES measures, no exclusions were made. All the ES measures are means from each participant over the whole period. RTs in the CN task were the dependent variable (long format) in the following analyses.

The basic language switching and mixing costs in the CN task were examined in a model with Language (L1 or L2) * Condition (repetition, switch, or single) as fixed factors, and random intercept for Participant and Stimulus.

In CN, there was a significant switching cost in L1 (E = 80.73, SE = 10.04, t = 8.04, p < .001) and a mixing cost (E = −73.01, SE = 7.14, t = −10.23, p < .001): switch trials were slower than repetition trials, and single block trials were faster than repetition trials. In L2, there was likewise a significant but slightly smaller switching cost (E = 50.37, SE = 10.26, t = 4.91, p < .001), and a mixing cost (E = −59.05, SE = 7.31, t = −8.08, p < .001). The difference in switch costs between languages was significant (E = 30.36, SE = 14.37, t = 2.11, p = .035), with the cost for L1 being larger.

Next, we examined the associations between the ES variables and the CN task with a model like the one above, but by including one of the ES variables at a time as a predictor. In L1, ES-IS did not predict either the switching or mixing cost (p's > .05), but in L2, ES-IS predicted the mixing cost (E = −73.33, SE = 18.56, t = −3.95, p < .001): the more the participant reported ES-IS, the higher their mixing cost was (Figure 2, panel A).

In L1, ES-US predicted both the switching cost (E = −83.29, SE = 36.92, t = −2.26, p = .024) and the mixing cost (E = -76.39, SE = 26.32, t = −2.90, p = .0037): the more US the participant reported, the higher their language mixing cost, but the smaller the switching cost (Figure 2, Panel B). In L2, ES-US predicted the mixing cost (E = −142.90, SE = 26.64, t = −5.37, p < .001): the more US, the larger the mixing cost (Figure 2, Panel C).

In L1, ES-WIS predicted the mixing cost (E = −48.17, SE = 20.53, t = −2.35, p = .019): the more WIS, the higher the mixing cost (Figure 2, Panel D). In L2, ES-WIS predicted both the mixing cost (E = −147.37, SE = 20.98, t = −7.03, p < .001) and the switch cost (E = −57.24, SE = 28.94, t = −1.98, p = .048). The mixing cost became larger the higher the WIS, and the switch cost became smaller (Figure 2, Panel E).

ES-BIS predicted neither the switch nor the mixing cost in either language (p's > .23).

3.2. Cued naming and the EF tasks

In the following models, RT in the CN task was the dependent variable in long format, whereas EF effects were included as means. The outlier exclusion in the CN task was done identically as in the previous analysis. In the executive tasks, a subject's mean score was deleted as an outlier if overall performance in that task failed to exceed chance level (.58); this resulted in removing data from one participant in the Flanker task. Additionally, any trial was excluded if it deviated more than +/−3SD from the participant's overall mean. This resulted in removing 36 trials in the Flanker task (1.13%), 45 trials in the Simon task (1.43%), and 70 trials in the Number-Letter task (1.52%).

To analyze the associations between the EF tasks and CN, we used a model with Language (L1 or L2) * Condition (repetition, switch, or single) as fixed factors, and random intercept for Participant and Stimulus, as well as one EF mean cost effect at a time as a predictor that was allowed to interact with Condition.

Simon effect predicted the CN switch cost in both L1 (E = −.17, SE = .047, t = −3.59, p < .001) and L2 (E = −.11, SE = .046, t = −2.48, p = .013): the higher the Simon effect, the smaller the switch cost (Figure 3, Panels A and B).[1]

Flanker effect predicted the CN switch cost in L1 (E = .12, SE = .031, t = 3.71, p < .001) and the mixing cost in L2 (E = −.085, SE = .022, t = −3.792, p < 001). The L1 switch cost was larger the higher the Flanker effect, and the L2 mixing cost was likewise larger the higher the Flanker effect (Figure 3, Panels C and D).

The Number-Letter switch cost did not predict the CN switch cost in either L1 or L2 (p's > .05). The Number-Letter mixing cost predicted the CN mixing cost in L1 (E = −.014, SE = .0044, t = −3.26, p = .0011), but not in L2 (p = .84). The L1 mixing cost was higher the larger the Number-Letter mixing cost (Figure 3, Panel E).[2]

In our previous study (Jylkkä et al., 2018b) using the same setup we found that a larger Simon effect predicted a smaller L2 switching cost; this finding was replicated in the current study. However, in the previous study we also found that the Simon effect predicted a smaller L1 mixing cost in the CN task, and that larger Number-Letter switching cost predicted a larger switching cost in the CN task; these effects were not replicated.

3.3. Everyday switching and the EF tasks

Associations between ES and the EF tasks were examined with lme-models with RT in the EF task as the dependent variable in long format (+/−3SD outliers removed), Condition as a fixed factor (congruent or incongruent in the Simon and Flanker tasks, and switch, repetition, or single in the Number-Letter task), and Participant and Trial as random factors. Additionally, one of the ES mean variables at a time was inserted as a continuous predictor that interacted with Condition.

None of the ES variables were associated with the Flanker effect (ES-IS p = .59; ES-US p = .12; ES-WIS p = .23), except for the ES-BIS interaction (E = 15.99, SE = 8.04, t = 1.99, p = .047), where higher frequency of BIS was associated with a larger Flanker effect (see Figure 4, Panel A). The Simon effect did not interact with the ES variables (ES-IS p = .70; ES-US p = .58; ES-WIS .37; ES-BIS p = .71).

As to the Number-Letter task, IS predicted both the mixing cost (E = −95.46, SE =22.43, t = −4.26, p < .001) and the switching cost (E = 79.03, E = 27.11, t = 2.92, p = .0036). Both the mixing

[1]Three of the participants had a negative Simon effect. We also conducted the analyses by excluding these participants. The direction of the results was identical with those reported above, with the exception that the CN switching cost in L2 was not significantly associated with the Simon effect.

[2]Two participants had a negative Number-Letter mixing cost, so we also conducted the analyses by excluding these participants. The results did not differ, with the exception that the association between the NL mixing cost and CN mixing cost in L1 was slightly weaker (E = −.011, SE = .0050, t = −2.22, p = .027).

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

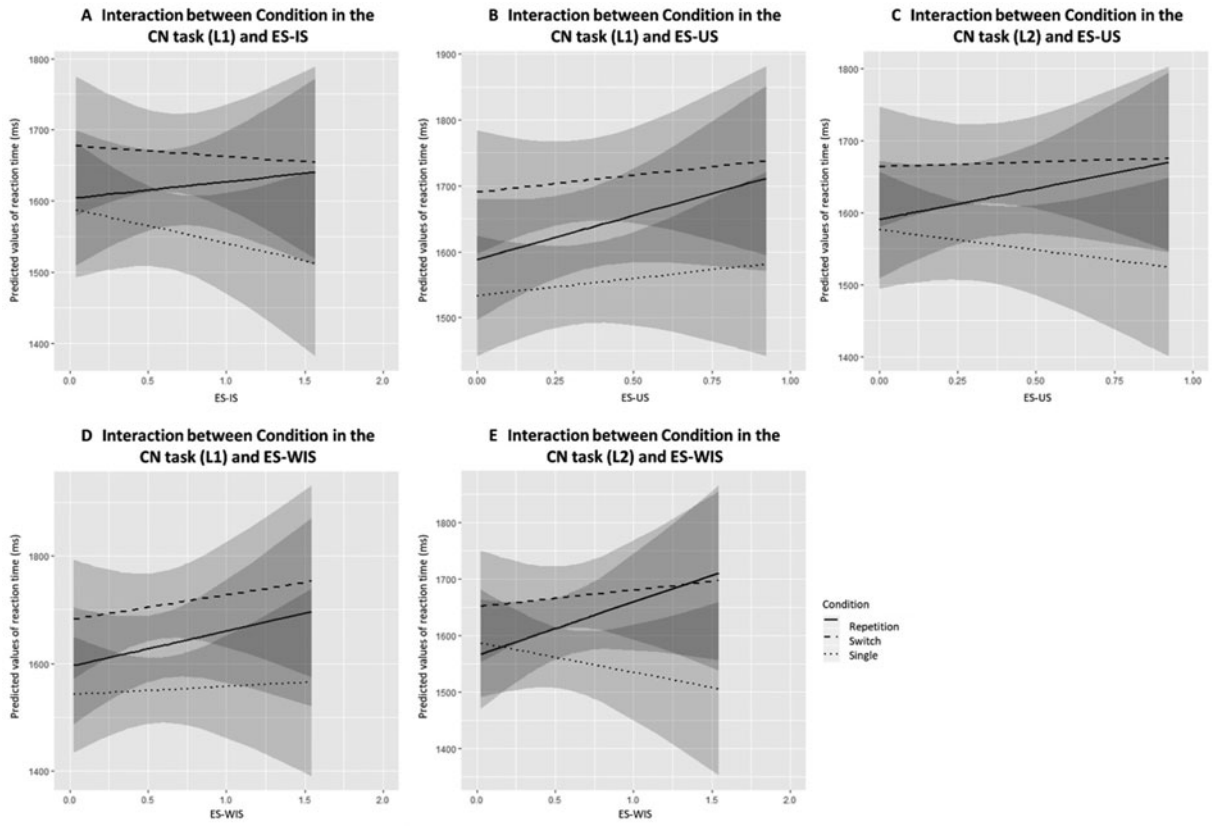


Fig. 2. Associations between switch and mixing costs in the Cued Naming (CN) task and the Everyday Switching (ES) variables. IS = Intended Switches; US = Unintended Switches; WIS = Within-Interlocutor Switches.

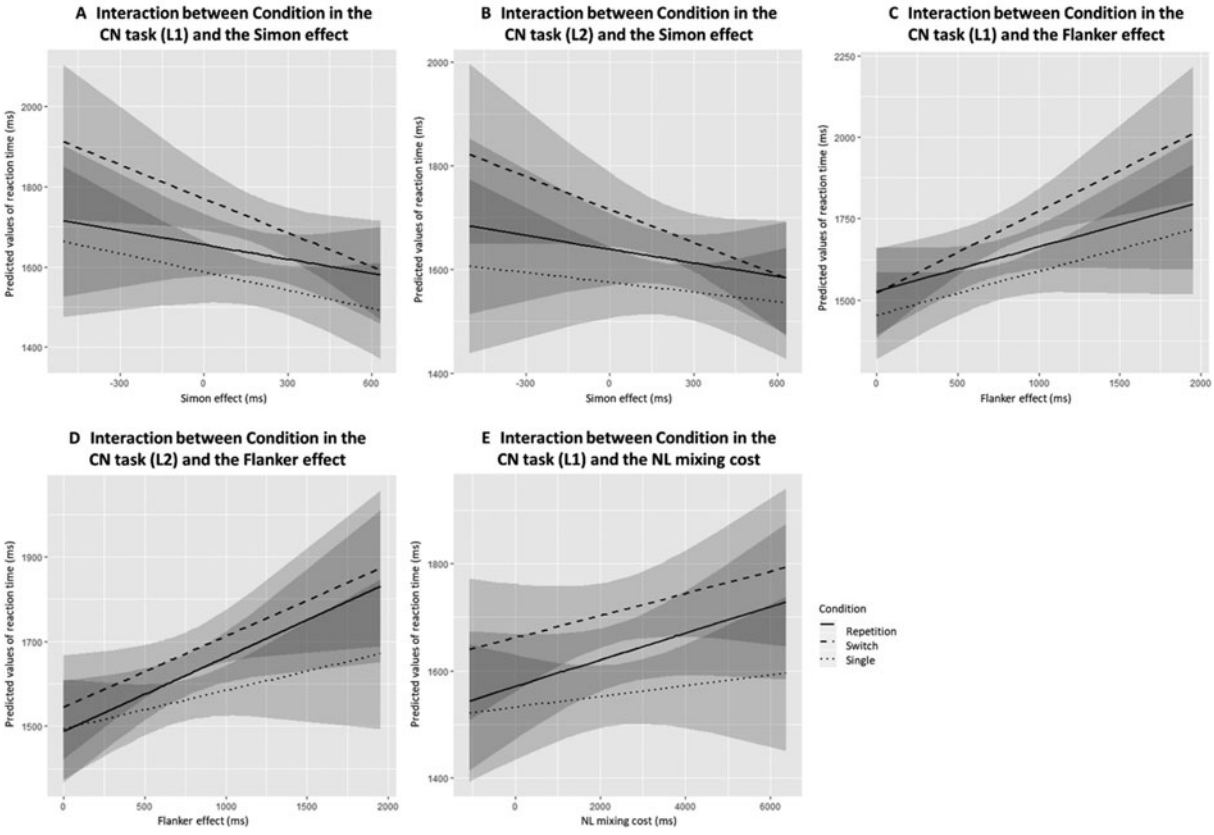


Fig. 3. Associations between the Cued Naming (CN) task and cost effects in the Executive Functions (EF) tasks. NL = Number-letter task.

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

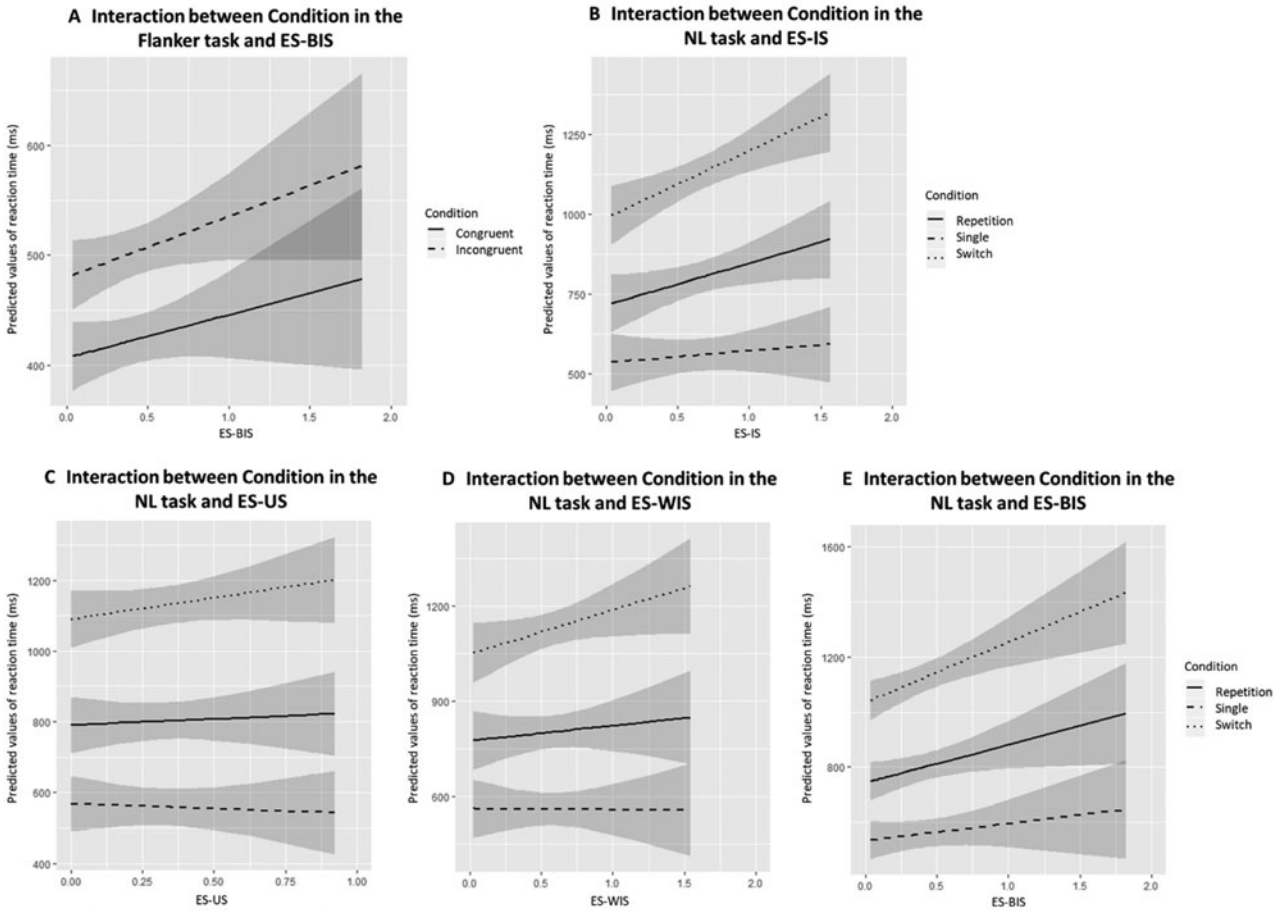


Fig. 4. Associations between the Everyday Switching (ES) variables and the executive task cost effects. NL = Number-letter task; IS = Intended Switches; US = Unintended Switches; WIS = Within-Interlocutor Switches; BIS = Between-Interlocutor Switches.

and switching cost were higher the more the participant reported IS (Figure 4, Panel B). US predicted the switch cost (E = 85.78, SE = 40.11, t = 2.14, p = .033) (higher frequency of US was associated with higher switch cost) but not the mixing cost (p = .060) (Figure 4, Panel C). WIS predicted both the switching cost (E = 90.51, SE = 30.53, t = 2.96, p = .0031) and the mixing cost (E = −50.98, SE = 25.34, t = −2.01, p = .044). The switching and mixing costs were higher the more the participant reported WIS (see Figure 4, Panel D). Finally, higher frequency of reported BIS predicted both higher mixing cost (−75.54, SE = 23.76, t = −3.18, p = .0015) and switching cost (E = 81.70, SE = 28.87, t = 2.83, p = .0047) (see Figure 4, Panel E).

In a previous study with a similar setup (Jylkkä et al., 2019), we found that a higher rate of Intended switches predicted a larger Number-Letter switching cost, and this finding was replicated in the present study. Moreover, in the previous study we found that a higher rate of Unintended switches predicted a smaller mixing cost in the Number-Letter task, but this finding was not replicated in the current study.

3.4. Correlations between the Simon and Flanker effects

The Simon and Flanker effects correlated negatively (r = −.35, p = .050). Correlations between Simon and Number-Letter or Flanker and Number-Letter were not examined, because they were hypothesized to tap on different functions (Miyake et al., 2000).

4. Discussion

We aimed to assess the associations between three types of behaviors in bilinguals that, according to the task-generality hypothesis, utilize overlapping cognitive mechanisms. These behaviors were everyday language switching, language switching in a cued picture naming task, and general computerized executive tests. If bilingual behaviors such as language switching could train general EF (e.g., Bialystok, 2009), these behaviors and EF test performances should rely on the same mechanisms and be correlated. As an alternative competing hypothesis, we have proposed the task-specificity hypothesis that is based on skill learning, where the cognitive system is assumed to adapt to specific tasks by building task-specific subroutines. If the skill learning approach were correct, then general EF would not be trainable through any specific behavior such as everyday language switching, at least if that behavior has already become routinized (Chein & Schneider, 2012; Paap, 2018). Accordingly, the presumably routinized everyday language switching behaviors in our bilinguals should not be associated with better EF performances. Again, the cued naming task, assuming that some of its aspects are relatively novel for the participants, would rely on general EF.

The results were, at least at face value, inconsistent and complex. In Table 2, we summarize the main findings to ease their interpretation, while estimates for all interactions are reported in Table 3. Based on the task-generality hypothesis, the expectations were as follows. (1) As to the associations between everyday

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

Table 2. Summary of the results. Statistically significant findings are reported as "Positive" or "Negative", depending on whether the correlation between the two measures is positive or negative. Findings that were in line with the task-generality hypothesis are marked with a smiley face (☺) and those that were against with a sad face (☹). Non-significant associations are denoted with "-", and non-applicable cells with "NA" (we did not examine associations between NL switching cost and Cued naming mixing cost, and between NL mixing cost and Cued naming switching cost).

		Cued naming (CN)		Executive tests (EF)		
	Switching cost	Mixing cost	Simon	Flanker	NL switching cost	NL mixing cost
Everyday switching frequency						
ES-IS	-	Positive ☹	-	-	Positive ☹	Positive ☹
ES-US	Negative ☹	Positive ☺	-	-	Positive ☺	-
ES-WIS	Negative	Positive	-	-	Positive	Positive
ES-BIS	-	-	-	Positive ☹	Positive ☹	Positive ☹
Executive tests						
Simon	Negative ☹	-				
Flanker	Positive ☺	Positive ☺				
NL switching cost	-	NA				
NL mixing cost	NA	Positive ☺				

NL = Number-Letter task; ES = Everyday Switching; IS = Intended Switches; US = Unintended Switches; WIS = Within-Interlocutor Switches; BIS = Between-Interlocutor Switches.

switching (ES) variables and cued naming (CN) measures, we expected that higher frequency of ES-IS would predict smaller CN switching and mixing costs, and that ES-US would predict larger CN switching and mixing costs. High rate of ES-IS can be taken to reflect flexible lexical control, whereas high frequency of ES-US can be taken to reflect lapses in lexical control (Jylkkä et al., 2017; Jylkkä et al., 2019; Rodriguez-Fornells et al., 2012). Following the Adaptive Control hypothesis (Green & Abutalebi, 2013), we expected that ES-WIS would reflect dense code switching and would not predict performance in the CN task, whereas ES-BIS would reflect dual language switching, and would be associated with better performance (smaller costs) in the CN task. (2) As to the associations between the executive tasks and the ES variables, we likewise expected that higher rates of ES-IS and ES-BIS would be associated with better executive performance (smaller EF cost effects), and that ES-US would predict worse executive performance, while ES-WIS would not be correlated with executive performance (because it is not assumed to engage EF). (3) Finally, in the CN and EF associations, we expected that overall better performance in the CN task would be associated with better performance in the EF tasks.

As can be seen from Table 2, less than half (5 out of 13) of the statistically significant findings were in line with the above hypotheses derived from the task-generality theory. We next examine the findings in more detail.

To start with the ES-CN associations, contrary to the task-generality theory, we found that ES, irrespective of type, was mainly associated with larger mixing costs in the CN task (see Tables 2 and 3). In contrast to five consistent and relatively strong effects in the case of CN mixing costs (see Table 3), we observed only two weak effects in the case of CN switching costs. Thus, we focus on the mixing costs. The CN mixing cost effect can be taken to reflect monitoring, or maintaining the simultaneous activation of two language task schemas in working memory. Even during repetition trials in the mixed block of the CN task, the participant is aware that the naming language can change at any moment, while in the single block only one task schema is active. Participants with higher CN mixing cost could be said to have more difficulty in maintaining two distinct language schemas in their working memory. Why is this associated with higher frequency of everyday language switching? It could be that, in participants with a high CN mixing cost, both languages are simultaneously more active than in participants with a low mixing cost, who succeed in directing activation to one language schema at a time. Such parallel activation of language sets could lead to higher frequency of language switching in everyday life. On this reasoning, we would expect to see more negative associations between language mixing and especially ES-US and ES-WIS, which hypothetically reflect unintentional switching or dense code switching, but not ES-IS or ES-BIS, which reflect more controlled language switching. No such dissociations could be observed. It is possible that the EMA questions failed to differentiate between these hypothetically distinct types of switches, or that the participants were unable to themselves categorize their switches into distinct types.

As regards the associations between the ES-measures and the executive tasks, seven out of eight interactions were significant in the case of the Number-Letter task, and in all cases higher frequency of everyday switching predicted larger NL switching and mixing costs (contrary to the task-generality hypothesis). The Simon and Flanker tasks, assumed to measure inhibition, were not associated with ES, let alone a significant association between the Flanker task and ES-BIS, where higher frequency of ES-BIS predicted a larger Flanker effect (again contrary to the task-generality hypothesis). Let us focus on the associations between ES and NL, because these were the most consistent. As was the case with the CN-ES associations, again we saw that a higher rate of everyday switching, irrespective of type, was associated with worse non-verbal monitoring capacity (as measured with the NL mixing cost), as well as worse non-verbal switching capacity (as measured with the NL switching cost). As above, we can reason that participants with a worse monitoring capacity have more difficulties in maintaining two distinct language sets simultaneously, so they also switch languages more frequently in everyday life.

The direction of the effect is contrary to what is usually assumed in bilingualism research: namely, that more frequent

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

Table 3. Estimates (standard error in parenthesis) for all interaction effects, including non-significant ones, separated by language in the Cued Naming task. Direction of the effects is based on visual inspection of the interaction plots and is depicted as follows: bold = positive (higher values on X predict larger effect); italic = negative (higher values on X predict smaller effect). Note that the sign of the estimate (+/-) alone does not indicate the direction of the interaction, because it has to be interpreted in relation to the intercept and main effects

	Cued naming (CN)				Executive tests (EF)			
	Switching cost		Mixing cost					
	L1	L2	L1	L2	Simon	Flanker	NL switching cost	NL mixing cost
Everyday switching frequency								
ES-IS	*−20.42 (25.14)*	*−34.73 (25.73)*	**−35.56 (18.28)†**	**−73.33 (18.56)*****	**2.88 (7.46)**	*−3.43 (6.28)*	**79.03 (27.11)****	**−95.46 (22.44)*****
ES-US	*−83.29 (36.92)**	*−66.07 (36.45)†*	**−76.39 (26.32)****	**−142.90 (26.64)*****	**5.91 (10.63)**	**12.72 (8.25)**	**85.78 (40.11)***	**−61.63 (32.81)†**
ES-WIS	*−18.56 (28.50)*	*−57.24 (28.94)**	**−48.17 (20.53)***	**−147.37 (20.98)*****	*−7.60 (8.43)*	**7.70 (6.48)**	**90.51 (30.53)****	**−50.98 (25.34)***
ES-BIS	*−23.09 (26.11)*	**1.04 (27.34)**	**−22.69 (19.06)**	**−6.13 (19.49)**	**2.94 (7.77)**	**15.99 (8.04)***	**81.70 (28.87)****	**−75.54 (23.76)****
Executive tests								
Simon*	*−.17 (.047)****	*−.11 (.046)**	**−.035 (.033)**	*.029 (.033)*				
Flanker	**.12 (.031)*****	*−.0099 (.031)*	*.0040 (.022)*	**−.085 (.022)*****				
NL switching cost	**.0014 (.0059)**	*−.011 (.0060)†*	NA	NA				
NL mixing cost*	NA	NA	**−.014 (.0044)****	**−.00092 (.0045)**				

†: p < .10; *: p < .05; **: p < .01; ***: p < .001.

language switching would be associated with better non-verbal switching ability. Here we observed the opposite: namely, better non-verbal switching ability was related to less frequent language switching. This indicates that frequent language switching cannot be hypothesized to train EF – it could be that people switch languages frequently exactly because they have low EF. It is unclear how this finding is related to the task-generality hypothesis. On the one hand, it could be argued that this finding supports the assumption that general EF are utilized in language switching, because otherwise there would be no significant associations whatsoever. On the other hand, the task-generality hypothesis specifically implies that better EF would be related to more fluent language switching.

As was the case with the ES-CN associations, the ES-EF effects did not differ between different types of reported language switching, which might indicate problems in the operationalization of the switching types, or simply that different types of language switches do not differ in how they are related to EF. If we assume that the ES-questions succeeded in tapping on different types of language switches, the fact that language switches irrespective of type were associated with worse EF performance speaks against the task-generality hypothesis, which specifically implies that more controlled language switches (IS and BIS) would be related to better EF.

In our previous study (Jylkkä et al., 2019) utilizing a similar setup, we found that a higher frequency of self-reported IS predicted a larger switching cost in the number-letter task, and that a higher frequency of US predicted a smaller mixing cost (everyday switching was measured with EMA also in the previous study). The first finding, but not the second, was replicated in the present study.

Finally, we observed associations between the EF tasks and the participants' performance in the CN task. In this case, the findings with respect to the mixing cost were roughly in line with our hypothesis: the Flanker effect and NL mixing costs were positively associated with the CN mixing cost: that is, worse general inhibitory or monitoring capacity predicted worse monitoring capacity in the CN task (2 significant findings out of 6; see Table 3). This can be taken to indicate that monitoring in the CN task might at least partly rely on the same type of monitoring processes that are utilized in non-verbal, general EF tasks. In an earlier study with a similar setup (Jylkkä et al., 2018b) we likewise found a positive association between the NL switching cost and the CN switching cost. This particular effect was not replicated here, but it is noteworthy that in both studies better CN performance was associated with better NL performance, which could be due to the structural similarity between these tasks.

In the present study, the Simon effect was negatively associated with the CN switching cost, a finding that is against our hypotheses: worse inhibitory capacity predicted better performance on the CN switching cost measure. However, the Flanker task was positively associated with the CN switching cost, in line with our hypotheses. In our previous study (Jylkkä et al., 2018b), we found, consistently with the present results, that larger Simon effect was associated with smaller CN switch cost, but also found a negative association with the CN mixing cost. Additionally, in the previous study the Flanker effect did not predict either the CN switch or the mixing cost, whereas here it was positively correlated with the CN switching cost. These discrepancies might indicate that the findings are not robust and/or that the tasks are not reliable. It is noteworthy that also previous studies (see section 1.2.) have been inconsistent in this regard. Thus, no

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

firm conclusions can currently be made about the relationship between cued bilingual picture naming and general EF.

We have thus far been discussing the results in light of the task-generality hypothesis. Let us now turn to the alternative explanation: the task-specificity hypothesis. This approach predicts that participants learn task-specific skills for routinized tasks, but rely on task-general mechanisms in novel tasks (Chein & Schneider, 2012). On this account, inter-task correlation should be proportional to how routinized the task-performance is. In line with this hypothesis, we observed positive correlations between the two more novel tasks, i.e., the CN and EF tasks. In turn, we found neg*ative* associations between the frequency of assumedly routinized everyday language switching and the novel CN and EF tasks. These negative associations could be fitted to the task-specificity hypothesis, if we assume that self-reported everyday switches largely reflect highly automatized language behaviors that are triggered by various contextual factors in a bottom-up fashion when executive control is weaker. This would not rule out the possibility of switching between language-schemas voluntarily and in a top-down manner when deemed as necessary, but such instances of language switching would be less common. On this account, the executive system has, as it were, a veto right to intervene in lower-level processes, which otherwise would run on autopilot (Baars, 1988; Dehaene & Changeux, 2011). Future research should attempt to test this speculative account, and better separate automatically driven vs. controlled language switches.

4.1. Summary of the findings

The present findings were mostly contrary to the task-generality hypothesis that better performance in the CN and EF tasks would be associated with higher rate of control-demanding language switches: more frequent everyday language switching, irrespective of type, was associated with worse performance in the cued naming task and the executive tasks. In the case of the EF-CN associations, the results were mostly consistent with the task-generality hypothesis: better CN performance was associated with better EF performance (Table 2). The overall pattern of the results is in line with the hypothesis about the strength of the associations depicted in Figure 1: we expected the associations to be more positive for EF-CN compared to the other domains. The reason for this is probably that the EF tasks are more similar to the CN task than ES is to any of the other tasks. Higher frequency of ES could reflect lack of language control, whereas control is arguably central in the laboratory tasks (CN and EF), where the response is always determined by the stimulus (not anything goes).

That some systematicity can be found in the results could be taken to support the task-generality hypothesis: everyday language switching frequency, language switching performance in the laboratory, and EF performance are somehow associated, although the associations are quite diverse and complex, at least in the present sample. The results do not, however, support the simple hypothesis that higher rate of language switching in everyday life would systematically be associated with better EF. In fact, the most consistent finding in the present study (i.e., the ES-EF associations) suggests the exact opposite: more frequent everyday language switching is associated with lack of control or monitoring capacity. This result could be accounted by the task-specificity account, assuming that a large part of everyday language switches relies on automatized submechanisms that are triggered easily, especially when executive control is weaker. On this account, top-down control of language switching is utilized only when dictated by one's goals. The task-specificity account is also in line with the associations between the EF and CN tasks: both include novel features and therefore rely on general executive mechanisms instead of task-specific subsystems.

An important aspect of the results is that the different ES variables, irrespective of type, mainly showed similar associations with the other variables, contrary to the predictions of the AC model. This can be explained in many different ways. It is possible that a) our operationalization of the everyday switching frequencies were not successful, or b) the participants could not themselves distinguish between different types of switches, or c) different types of language switches are not differently associated with EF and CN, or d) there simply are no distinguishable language switch types. Again, if we assume that the ES-BIS questions did manage to tap on between-interlocutor language switches, then the significant findings were exactly contrary to the predictions of the AC hypothesis. Higher rate of ES-BIS was associated with worse executive performance, whereas the AC hypothesis claims that they require cognitive control.

A central problem in studies on executive functions is the task impurity problem: tasks that hypothetically tap on the same executive mechanism often fail to correlate with each other (Miyake et al., 2000; Paap et al., 2015; Paap & Sawi, 2014). In the present study, we only had two tests that hypothetically measured the same functions: namely, Simon and Flanker tasks on inhibition. The two tasks showed a negative correlation, which speaks against the assumption that they would tap a common, general inhibitory mechanism. To compare, in Jylkkä et al. (2019) these two tasks were not significantly correlated ($r = .24$, $p = .19$, $N = 33$); in Jylkkä et al. (2017) there was a small positive correlation (Experiment 1: $r = .21$, $p = .098$, $N = 66$; Experiment 2: $r = .22$, $p = .025$, $N = 112$); and in Jylkkä et al. (2018b) there was no correlation ($r = .060$, $p = .68$, $N = 51$). These findings are not surprising, given that even two versions of the same task often show low or non-significant correlations (cf. Salthouse, 2010; Shilling et al., 2002). The take-home message is that if executive tasks that are structurally similar fail to show robust correlations, we cannot expect them to be consistently correlated with highly different behaviors, such as bilingual language switching. A skill-learning explanation for the lack of inter-task correlations could be that participants quickly learn task-specific subroutines, thereby disengaging general EF when the task becomes familiar.

4.2. Limitations and strengths

The main limitation of the present study is its relatively small sample size. The main reason for this is that the EMA method is time-consuming and requires commitment from the participants. However, the use of multilevel models partially accommodates for small sample size, because it maximizes statistical power through analyzing all observations instead of just overall means. In addition to sample size, another limitation is that the tested sample had a relatively low language switching frequency and there were no large differences in different types of language switches they made. Thus, the results might not be generalizable to other populations, such as those where language switching is more frequent. Future studies should investigate the associations in other kinds of bilinguals who are functioning in environments with different code-switching and mixing practices than the ones in the present study. The strength of the present study is that it is

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

the first to examine associations between three different types of behaviors, which all hypothetically rely on similar cognitive mechanisms.

5. Conclusion

The possibility of enhancement of domain-general executive functions through behaviors such as language switching requires that such behaviors engage task-general executive functions. In other words, they require that the task-generality hypothesis is true. The results of the present study indicate that higher rate of everyday language switching is not associated with better executive functions, but instead quite the opposite: participants with lower executive functioning tended to switch more often in our sample. This conflicts with the bilingual training hypothesis, which implies that language switching can train executive functions. The results were somewhat in line with the task-specificity hypothesis which implies that routinized, everyday language switching is an automatized process that does not require top-down control. The findings concur with previous studies, which have likewise failed to discover systematic associations between language switching and executive functions. This could be the reason why the recent extensive meta-analysis (Lehtonen et al., 2018) did not find an executive advantage in bilinguals.

Acknowledgements. We thank Cristel Merilä and Juhani Virta for help in testing the participants and coding the data. Minna Lehtonen was supported by Academy of Finland (#288880) and Matti Laine and Jussi Jylkkä by the Abo Akademi University Endowment (the BrainTrain project and a personal grant).

References

Baars, BJ (1988) *A cognitive theory of consciousness*, Cambridge: Cambridge UP.

Bialystok, E (2009) Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition* **12**(1), 3–11. https://doi.org/10.1017/S1366728908003477

Botvinick, MM, Braver, TS, Barch, DM, Carter, CS and Cohen, JD (2001) Conflict monitoring and cognitive control. *Psychological Review* **108**(3), 624–652. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11488380

Calabria, M, Branzi, FM, Marne, P, Hernández, M and Costa, A (2015) Age-related effects over bilingual language control and executive control. *Bilingualism: Language and Cognition* **18**(1), 65–78. https://doi.org/10.1017/S1366728913000138

Calabria, M, Hernández, M, Branzi, FM and Costa, A (2012) Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology* **3**(JAN), 399. https://doi.org/10.3389/fpsyg.2011.00399

Chein, JM and Schneider, W (2012) The Brain's Learning and Control Architecture. *Current Directions in Psychological Science* **21**(2), 78–84. https://doi.org/10.1177/0963721411434977

Declerck, M, Grainger, J, Koch, I and Philipp, AM (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. https://doi.org/10.1016/j.jml.2017.03.005

Dehaene, S and Changeux, JP (2011) Experimental and theoretical approaches to conscious processing. *Neuron* **70**(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Dick, AS, Garcia, NL, Pruden, SM, Thompson, WK, Hawes, SW, Sutherland, MT, Riedel, MC, Laird, AR and Gonzalez, R (2019) No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nature Human Behaviour* **3**(7), 692. https://doi.org/10.1038/s41562-019-0609-3

Eriksen, BA and Eriksen, CW (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* **16**(1), 143–149. https://doi.org/10.3758/BF03203267

Gathercole, SE, Dunning, DL, Holmes, J and Norris, D (2019) Working memory training involves learning new skills. *Journal of Memory and Language* **105**, 19–42. https://doi.org/10.1016/J.JML.2018.10.003

Green, DW (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* **1**(2), 67–81. https://doi.org/10.1017/S1366728998000133

Green, DW and Abutalebi, J (2013) Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology* **25**(5), 1–16. https://doi.org/10.1080/20445911.2013.796377

Hartanto, A and Yang, H (2016) Disparate bilingual experiences modulate task-switching advantages: A diffusion-model analysis of the effects of interactional context on switch costs. *Cognition* **150**, 10–19. https://doi.org/10.1016/j.cognition.2016.01.016

Johnson, HA, Sawi, O and Paap, KR (2015) Language switching frequency in bilinguals is inconsistently linked to executive functioning, In *Annual Meeting of the Cognitive Neuroscience Society*.

Jylkkä, J (2017) *Bilingual Language Switching and Executive Functions*, Åbo Akademi. Retrieved from http://www.doria.fi/bitstream/handle/10024/147587/jylkka_jussi.pdf?sequence=2

Jylkkä, J, Soveri, A, Wahlström, J, Lehtonen, M, Rodríguez-Fornells, A and Laine, M (2017) Relationship between language switching experience and executive functions in bilinguals: an Internet-based study. *Journal of Cognitive Psychology* **29**(4), 1–16. https://doi.org/10.1080/20445911.2017.1282489

Jylkkä, J, Lehtonen, M, Kuusakoski, A, Lindholm, F, Hut, SCA and Laine, M (2018a) The role of general executive functions in receptive language switching and monitoring. *Bilingualism: Language and Cognition* **21**(4), 1–17. https://doi.org/10.1017/S1366728917000384

Jylkkä, J, Lehtonen, M, Lindholm, F, Kuusakoski, A and Laine, M (2018b) The relationship between general executive functions and bilingual switching and monitoring in language production. *Bilingualism: Language and Cognition* **21**(3). https://doi.org/10.1017/S1366728917000104

Jylkkä, J, Soveri, A, Laine, M and Lehtonen, M (2019) Assessing bilingual language switching behavior with ecological momentary assessment. *Bilingualism: Language and Cognition* **23** (2). https://doi.org/10.1017/S1366728918001190

Koch, I, Prinz, W and Allport, A (2005) Involuntary retrieval in alphabet-arithmetic tasks: Task-mixing and task-switching costs. *Psychological Research* **69**(4), 252–261. https://doi.org/10.1007/s00426-004-0180-y

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018) Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin* **144**(4), 394–425. https://doi.org/10.1037/bul0000142

Linck, JA, Schwieter, JW and Sunderman, G (2012) Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition* **15**(03), 651–662. https://doi.org/doi:10.1017/S136672891100054X

Meuter, RFI and Allport, A (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* **40**(1), 25–40.

Miyake, A, Friedman, NP, Emerson, MJ, Witzki, AH, Howerter, A and Wager, TD (2000) The unity and diversity of executive functions and their contributions to complex Frontal Lobe tasks: a latent variable analysis. *Cognitive Psychology* **41**(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Paap, KR (2018) Bilingualism in Cognitive Science. In *The Cambridge Handbook of Bilingualism* (pp. 435–465). Cambridge: Cambridge UP.

Paap, KR, Johnson, HA and Sawi, O (2015) Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* **69**, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014

Paap, KR, Myuz, HA, Anders, RT, Bockelman, MF, Mikulinsky, R and Sawi, OM (2017) No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology* **29**(2), 89–112. https://doi.org/10.1080/20445911.2016.1248436

Paap, KR and Sawi, O (2014) Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press

identification of the theoretical constructs. *Frontiers in Psychology* **5**(962), 1–15. https://doi.org/10.3389/fpsyg.2014.00962

Prior, A and Gollan, TH (2011) Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society* **17**(4), 682–691. https://doi.org/10.1017/S1355617711000580

Prior, A and MacWhinney, B (2010) A bilingual advantage in task switching. *Bilingualism* **13**(2), 253–262. https://doi.org/10.1017/S1366728909990526

Rodriguez-Fornells, A, De Diego Balaguer, R and Münte, TF (2006) Executive control in bilingual language processing. *Language Learning* **56** (s1), 133–190. https://doi.org/10.1111/j.1467-9922.2006.00359.x

Rodriguez-Fornells, A, Krämer, UM, Lorenzo-Seva, U, Festman, J and Münte, TF (2012) Self-assessment of individual differences in language switching *Frontiers in Psychology* **3**(388), 1–15. https://doi.org/10.3389/fpsyg.2011.00388

Rogers, RD and Monsell, S (1995) Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General* **124**(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Rubin, O and 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**(6), 1477–1491. https://doi.org/10.1037/0278-7393.31.6.1477

Salthouse, TA (2010) Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain and Cognition* **73**(1), 51–61. https://doi.org/10.1016/j.bandc.2010.02.003

Segal, D, Stasenko, A and Gollan, TH (2019) More evidence that a switch is not (always) a switch: Binning bilinguals reveals dissociations between task and language switching. *Journal of Experimental Psychology: General* **148** (3), 501–519. https://doi.org/10.1037/xge0000515

Shilling, VM, Chetwynd, A and Rabbitt, PMA (2002) Individual inconsistency across measures of inhibition: an investigation of the construct validity of inhibition in older adults. *Neuropsychologia* **40**(6), 605–619. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11792402

Simon, JR and Rudell, AP (1967) Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *The Journal of Applied Psychology* **51**(3), 300–304. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6045637

Soveri, A, Antfolk, J, Karlsson, L, Salo, B and Laine, M (2017) Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-016-1217-0

Soveri, A, Rodriguez-Fornells, A and Laine, M (2011) Is there a relationship between language switching and executive functions in bilingualism? Introducing a within-group analysis approach. *Frontiers in Psychology* **2** (183), 1–8. https://doi.org/10.3389/fpsyg.2011.00183

Stasenko, A, Matt, GE and Gollan, TH (2017) A relative bilingual advantage in switching with preparation: Nuanced explorations of the proposed association between bilingualism and task switching. *Journal of Experimental Psychology: General* **146**(11), 1527–1550. https://doi.org/10.1037/xge0000340

Struys, E., Woumans, E., Nour, S., Kepinska, O. and Van Den Noort, M (2019) A domain-general monitoring account of language switching in recognition tasks: Evidence for adaptive control. *Bilingualism* **22**(3), 606–623. https://doi.org/10.1017/S1366728918000342

Taatgen, NA (2013) The nature and transfer of cognitive skills. *Psychological Review* **120**(3), 439–471. https://doi.org/10.1037/a0033138

Timmer, K, Calabria, M and Costa, A (2019) Non-linguistic effects of language switching training. *Cognition* **182**, 14–24. https://doi.org/10.1016/j.cognition.2018.09.001

Verreyt, N, Woumans, E, Vandelanotte, D, Szmalec, A and Duyck, W (2016) The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition* **19**(1), 1–10. https://doi.org/10.1017/S1366728914000352

Yim, O and Bialystok, E (2012) Degree of conversational code-switching enhances verbal task switching in Cantonese–English bilinguals. *Bilingualism: Language and Cognition* **15**(04), 873–883. https://doi.org/10.1017/S1366728912000478

https://doi.org/10.1017/S1366728920000619 Published online by Cambridge University Press