

Jussi Jylkkä

# Bilingual Language Switching and Executive Functions





# Bilingual Language Switching and Executive Functions

Jussi Jylkkä

Psychology  
Department of Psychology and Logopedics  
Åbo Akademi University  
Åbo, Finland, 2017

**Supervised by**

Docent Minna Lehtonen, PhD  
Department of Psychology  
Åbo Akademi University

Professor Matti Laine, PhD  
Department of Psychology  
Åbo Akademi University

**Reviewed by**

Professor Raymond Klein, Dalhousie University, Canada  
Professor Andrea M. Philipp, Aachen University, Germany

**Opponent**

Kenneth R. Paap, San Francisco State University, USA

ISBN 978-952-12-3627-3 (Print)

ISBN 978-952-12-3628-0 (Pdf)

Painosalama, Turku, Finland 2017

# Table of contents

Acknowledgments.....	4
List of original publications.....	5
Svensk sammanfattning.....	6
Abstract.....	7
1. Introduction.....	8
1.1. Earlier research on the Engagement and Training hypotheses.....	10
1.1.1. Studies on the Engagement Hypothesis.....	10
1.1.2. Studies addressing the Training Hypothesis.....	13
1.1.3. Summary of earlier research.....	14
1.2. The present study.....	14
1.3. Aims and hypotheses.....	16
2. Method.....	17
2.1. Participants.....	17
2.2. Procedure.....	17
2.2.1. The executive tasks.....	17
2.2.2. Everyday language switching behavior.....	18
2.2.3. Language switching laboratory tasks.....	21
3. Results.....	22
3.1. Everyday language switching and executive performance.....	22
3.1.1. Reliability and validity of General Language Switching Questions.....	22
3.2. Language switching in laboratory and executive performance.....	25
4. Discussion of the individual studies.....	27
4.1. Everyday language switching and executive performance.....	27
4.2. Language switching in laboratory and executive performance.....	28
4.3. Inhibition and asymmetry.....	31
5. General discussion.....	32
5.1. Within-lexicon control processes.....	33
5.2. Task-specific control processes.....	33
5.3. Directions for future research.....	37
6. Conclusions.....	40
References.....	41
Original articles.....	47

## Acknowledgments

First and foremost, I am deeply indebted to my supervisors Minna Lehtonen and Matti Laine, without whom this project would not have been possible. This is for two reasons. First, had I not been hired to work on their project on bilingualism and executive functions, I would never have started writing this doctoral thesis in the first place. Second, and more importantly, they have always been available for extremely helpful discussions regarding any questions or comments I have had. I believe we have learnt a lot from each other during these three years, and it has been a privilege to get acquainted with you both.

I would like to thank all members of the BrainTrain research group, whose seminars have provided a warm and stimulating forum to discuss both our own and each other's research, as well as broader scientific, ethical, or even philosophical questions. I am particularly grateful to Anna Soveri for her invaluable help in polishing the Study 2 manuscript.

I thank Henri Olkonieniemi for his guidance with statistical analyses, Juhani Virta for his help with data gathering and coding, and Teemu Laine for help with programming. I am also indebted to our excellent master students Anna Kuusakoski and Fred Lindholm for their great effort in gathering data for Studies 3 and 4.

I would like to express my deepest gratitude to Professor Raymond Klein and Professor Andrea Philipp for their rigorous examination of this thesis, as well as their kind and insightful comments. I also wish to express my sincere appreciation to Professor Kenneth Paap, who kindly accepted the invitation to travel from San Francisco to cold and rainy Åbo to discuss bilingualism and executive functions.

Finally, I would like to thank my family for their love, support, and patience.

Åbo, November 2017

The image shows a handwritten signature in black ink. The signature is stylized and appears to be 'Jussi Jylkkä'. It consists of a large, sweeping initial 'J' followed by a series of loops and a long horizontal tail.

Jussi Jylkkä

## List of original publications

1. 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, 404-419, DOI: 10.1080/20445911.2017.1282489
2. Jylkkä, J., Soveri, A., Laine, M., Lehtonen, M. (submitted). Assessing bilingual language switching behavior with Ecological Momentary Assessment. Manuscript submitted for publication.
3. Jylkkä, J., Lehtonen, M., Lindholm, F., Kuusakoski, A., and Laine, M. (in press). The relationship between general executive functions and bilingual switching and monitoring in language production. *Bilingualism: Language and Cognition*. DOI: <https://doi.org/10.1017/S1366728917000104>
4. Jylkkä, J., Lehtonen, M., Kuusakoski, A., Lindholm, F., Hut, S. A. H., and Laine, M. (in press). Receptive language switching and monitoring: The role of general executive functions in receptive language switching and monitoring. *Bilingualism: Language and Cognition*. DOI: <https://doi.org/10.1017/S1366728917000384>

The original articles have been reproduced by permission of the copyright holders.

## Svensk sammanfattning

Tvåspråkigas exekutiva förmåga har förslagits vara bättre än enspråkigas (exekutiva försprånget hos tvåspråkiga -hypotesen; EFT). Man har antagit att denna fördel uppstår som en följd av att tvåspråkiga individer kontinuerligt tränar sina exekutiva funktioner (EF) på ett naturligt vis genom att byta språk eller inhibera det språk som inte för tillfället används. Nyligen har forskning dock ifrågasatt den EFT-hypotesen. Syftet med föreliggande avhandling var att undersöka vilka de teoretiska antagandena som ligger till grund för EFT är och ifall dessa antaganden fylls. En skillnad görs mellan en hypotes, enligt vilken språkbytesbeteende hos tvåspråkiga engagerar allmänna exekutiva funktioner (engageringshypotes, EH) och en hypotes, enligt vilken tvåspråkigas verbala beteende tränar exekutiva funktioner (träningshypotes, TH). I föreliggande avhandling presenteras fyra studier, i vilka det å ena sidan undersöks ifall högre vardaglig språkbytesfrekvens hos tvåspråkiga har samband med deras exekutiva prestation, och å andra sidan undersöks det ifall tvåspråkigas exekutiva förmåga predicerar deras språkbytest prestation i laboratoriet. TH predicerar att de tvåspråkiga som byter mera i deras vardagliga liv har fått mera träning i EF och därför presterar bättre i test som mäter EF. Samtidigt implicerar UH att personer som har högre EF också är bättre i att byta mellan språk. Resultaten stödde ingendera av dessa hypoteser: den vardagliga bytesfrekvensen var inte associerad med bättre EF och det fanns inga konsistenta samband mellan språkbytest prestation i laboratoriet och EF. Fynden tyder på att de antaganden som ligger till grund för EFT är felaktiga. Däremot är resultaten konsistenta med hypotesen, enligt vilken den underliggande mekanismerna för språkbyte är specifika i stället för generella, dvs. inte delas med andra kognitiva uppgifter.

## Abstract

According to the Bilingual Executive Advantage (BEA) hypothesis, bilinguals outperform monolinguals in executive functions (EF). This has been suggested to stem from life-long training of bilinguals in shifting between languages and inhibiting the non-target language that are assumed to load on executive resources. Recent research, however, has brought the BEA hypothesis into doubt. In this thesis, I examine what the prerequisites are for bilingual training of EF to occur, and whether they are fulfilled. This is motivated by the fact that these assumptions are rarely examined. I distinguish between the Engagement and Training Hypotheses. The former holds that bilingual language use utilizes general EF; the latter implies that general EF are enhanced through training. Both hypotheses are necessary for BEA to occur. The Engagement Hypothesis entails that a bilingual person's language switching and general task switching or inhibition are inherently linked as they tap the same general executive resources. The Training Hypothesis implies that these executive resources can be enhanced through practice. I present four studies that examine, first, whether higher frequency of self-reported everyday language switching is associated with better EF in bilinguals and second, whether bilinguals' language switching and mixing performances in laboratory are associated with their general executive performance. The results indicated that in contrast to the Training Hypothesis, higher switching frequency was not associated with better EF. Moreover, in contrast to the Engagement Hypothesis, there were no consistent associations between bilinguals' language switching and mixing performance in the laboratory tasks and their general EF capacity. All in all, the results suggest that the underlying assumptions of the BEA hypothesis are incorrect. Instead, the present findings indicate that the mechanisms underlying language switching and mixing in bilinguals are largely task-specific and not based on general executive resources.



# 1. Introduction

Executive functions (EF) are cognitive control mechanisms that are central for a successful completion of tasks, and they include cognitive flexibility, inhibition, and working memory (Miyake et al., 2000). There is evidence that, at least in children, EF are more important for academic achievement than intelligence quotient (Blair & Razza, 2007). It is unclear to what extent executive performance can be shaped by environment and to what extent it is genetic in origin (e.g. Friedman et al., 2009; Karbach & Kray, 2009; Karbach & Verhaeghen, 2014). Given the central role of EF in everyday life and academic achievement, training programs to enhance executive performance have been developed (e.g. Karbach & Verhaeghen, 2014). Also, bilingual language use has been suggested to train EF (e.g. Bialystok, 2017). For instance, switching between languages is assumed to train general set-shifting and monitoring, and avoiding intrusion from the dominant language into the weaker language is assumed to enhance general inhibition (e.g. Linck, Schwieter, & Sunderman, 2012; Rodriguez-Fornells, De Diego Balaguer, & Münte, 2006). However, recent meta-analyses and reviews of the bilingual executive advantage (BEA) compared to monolinguals indicate that the advantage is negligible or non-existent (e.g. Hilchey & Klein, 2011; Lehtonen et al., 2017; Paap, Johnson, & Sawi, 2015), and that there is also a publication or reporting bias favoring studies with positive results (de Bruin, Treccani, & Della Sala, 2015). This motivates examining the BEA and its underlying assumptions further.

In this study, I focus on the mechanisms underlying the hypothesized bilingual executive advantage. Here it is important to distinguish between two distinct assumptions of the BEA, namely the Engagement Hypothesis (EH) and the Training Hypothesis (TH):

(EH) Bilingual linguistic behaviors *engage* domain-general executive functions

(TH) Bilingual linguistic behaviors *train* domain-general executive functions

The latter presupposes the former: if bilingual behaviors did not engage domain-general executive functions, they could not possibly train them. Moreover, it is possible that simply loading on a cognitive system might not enhance its performance (that is, TH might not be true even if EH was).

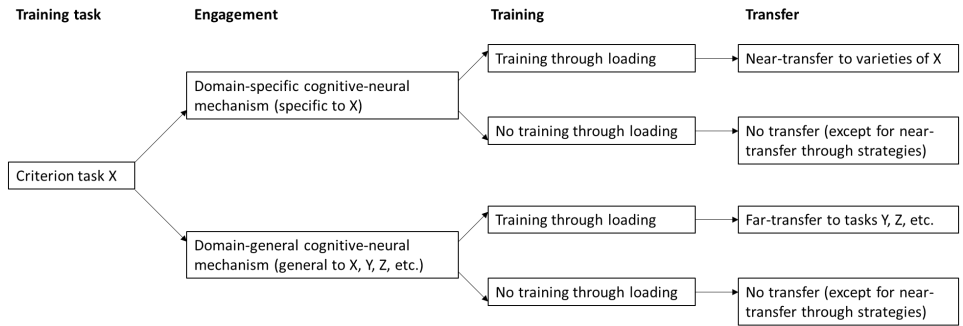
These questions are closely related to research on cognitive enhancement generally, which examines the possibilities to train working memory or executive performance through adaptive training tasks. Typically this approach endorses a “muscle metaphor” of the brain, namely that exposing the brain to increasing loads

at the capacity limits forces the brain to adapt through neuroplasticity, which results in enhanced performance in the task that is trained (the “criterion task”) and other tasks that rely on the same cognitive-neural mechanisms (e.g. Klingberg, 2010). If training on the criterion task enhances performance within the same cognitive domain, “near-transfer” is observed, whereas if the criterion task trains tasks that are structurally different from the criterion task, one has succeeded in eliciting “far-transfer” (e.g. Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Jaeggi, Buschkuhl, Shah, & Jonides, 2014). Cognitive training typically aims at far transfer, that is, generalized effects of training in academic success or coping with everyday challenges, not merely enhancement in untrained laboratory tasks that represent the same domain as the criterion task.

In the case of bilingual linguistic behaviors, “criterion tasks” can be taken to include language switching and monitoring. The more bilinguals switch between languages, the better they should become at that task. The key question is, however, whether language switching experience enhances performance in other tasks that hypothetically engage the same mechanisms, such as general inhibitory or set-shifting capacity. In other words, of importance is whether bilingual language use elicits far transfer to non-linguistic executive performances.

Figure 1 summarizes the possible routes from training on a criterion task (such as bilingual language switching) to types of transfer in other tasks. The engagement of domain-general cognitive mechanisms is a necessary but not sufficient condition for far transfer to occur: it is possible that loading on a cognitive mechanism might not result in adaptive changes in the mechanism. Even if training resulted in enhancement in the criterion task and the task engaged domain-general functions, the enhancement in the criterion task could be due to use of strategies (e.g. chunking in the case of working memory) instead of improvement in the mechanism itself (cf. Dunlosky & Kane, 2007; Kaakinen & Hyönä, 2007; McNamara & Scott, 2001). Strategies are typically task-specific and would thus only result in near transfer. Far transfer arguably occurs only when the criterion task both engages and trains a domain-general cognitive-neural mechanism.

Figure 1. A flowchart of the routes possibly leading to transfer from a criterion task X.



## 1.1. Earlier research on the Engagement and Training hypotheses

I will next examine previous studies addressing the Engagement and Training hypotheses. Studies addressing EH include both brain imaging and behavioral studies, and have typically examined associations between general EF and linguistic tasks within bilingual samples, focusing on whether language switching and general EF tasks engage the same anatomical brain regions, or whether there are associations in a subject's performance between the two types of tasks. The studies addressing TH, in turn, have typically relied on between-groups designs, comparing the executive performance of bilinguals and monolinguals: if bilingual experience trains general EF, then bilinguals should outperform monolinguals on EF tasks. However, studies on TH also include within-group designs, which have examined whether the everyday language switching frequency of bilinguals is associated with their general EF performance. TH implies that the more a bilingual switches between languages, the higher their performance in general executive tasks should be.

### 1.1.1. Studies on the Engagement Hypothesis

Functional magnetic resonance imaging (fMRI) studies indicate that bilingual language control engages areas involved in general executive functions. The executive system is typically taken to mainly involve areas in the prefrontal and parietal cortices, in particular the anterior cingulate cortex (ACC) and dorsolateral prefrontal cortex (DLPFC), which are considered to be central in conflict detection and resolution (e.g. Nee, Wager, & Jonides, 2007; Niendam et al., 2012). Several studies indicate that the same network is active in bilinguals during tasks that require language control. For example, Van Heuven, Schriefers, Dijkstra, and Hagoort (2008) found that the lateral intraparietal cortex (LIPC) and anterior

cingulate cortex (ACC) were associated with between-language conflict in a lexical decision task in bilinguals, which can be taken to suggest that the general executive system is involved in suppressing the non-target language in the task. There is also evidence that areas of the executive network are involved in language switching, including areas such as ACC, DLPFC, and left inferior frontal gyrus (LIFG; e.g. Abutalebi et al., 2007; Abutalebi & Green, 2008; Crinion et al., 2006; Guo, Liu, Misra, & Kroll, 2011; Hernandez, Martinez, & Kohnert, 2000; Price, Green, & Von Studnitz, 1999; Wang, Kuhl, Chen, & Dong, 2009).

These brain imaging studies support the Engagement Hypothesis, albeit only indirectly. Even though general executive tasks and language switching rely on the same anatomical regions, this does not guarantee that they rely on the same functions, given that multiple functions can be realized in the same relatively large-scale structures that current fMRI methods can discern. In this respect, behavioral data can be considered as more direct evidence of whether language switching relies on the same cognitive functions as general executive tasks.

There are surprisingly few behavioral studies addressing the relationship between general EF and language switching. Linck et al. (2012) examined the relationship between a subject's performance in the Simon task, assumed to tap on general inhibition, and their language switching capacity in a picture naming task. They found that subjects with better performance in the Simon task also performed better in the language switching task, in line with the Engagement Hypothesis. The weakness of the Linck et al. study is their stimuli; they utilized only 10 black-and-white drawings which were repeated 60 times during the task. This type of setup, wherein stimulus is presented repeatedly, resembles a rapid automatized naming (RAN) task. As argued by Bexkens, Van Den Wildenberg, and Tijms (2015), in RAN the subjects may store the stimulus-name mappings in short-term memory in a highly accessible state. According to Bexkens et al. (2015), the competition between the current and previous stimuli in this type of task is resolved using inhibition. Thus, it is possible that the experimental setup of Linck et al. was particularly taxing on inhibition.

Similarly to the Linck et al. study, Liu, Rossi, Zhou, and Chen (2014) examined the language switch costs in a picture naming task between subjects with high vs. low inhibitory control (IC), measured with a Simon task. They found symmetric language switch costs between L1 and L2 in subjects with high IC, and asymmetric language switch costs (larger cost when switching into L1 than L2) in subjects with low IC. This type of language switch cost asymmetry is typically taken to reflect the need to inhibit the stronger language L1 more than the weaker L2. Hypothetically, when switching from L2 to L1, the strong L1 inhibition present in the previous L2 trial has to be resolved, leading to larger costs when switching into L1 than when

switching into L2 (Green, 1998). Thus, the results of Liu et al (2014) can be considered to support the Engagement Hypothesis. In another study by the same group, Liu, Fan, Rossi, Yao, and Chen (2016) examined the associations between cognitive flexibility (CF), assessed with Wisconsin Card Sorting Test (WCST) and language switch costs. In line with their earlier results, they found symmetrical language switch costs in the high CF group, but asymmetric switch costs in the low CF group. A weakness of both these studies is that they used the Simon performance as a dichotomous variable, lessening the power of the setup.

Against the Engagement Hypothesis are dissociations in bilinguals' performance between linguistic and non-linguistic executive tasks. For example, Calabria, Hernández, Branzi, and Costa (2012) examined bilinguals' performance in a cued naming task and a non-linguistic card-sorting test, and found that the switch costs were symmetrical in the former and asymmetrical in the latter. If the same general control mechanism were utilized by both tasks, we would expect similar switch cost asymmetries. Similarly, Calabria, Branzi, Marne, Hernández, and Costa (2013) examined age-related changes in a linguistic and non-linguistic switching task in bilinguals, and found an age-related change only in the non-linguistic task (see also Magezi, Khateb, Mouthon, Spierer, & Annoni, 2012; Weissberger, Wierenga, Bondi, & Gollan, 2012).

Similar dissociations between linguistic and non-linguistic tasks have been found also in patients with aphasia. For example, Green et al. (2010) examined two aphasic patients, both equally impaired in both languages. The first patient had subcortical lesions and the other abnormalities in left frontal and temporal, and bilateral occipital regions. The first patient manifested abnormal Stroop performance but less abnormal Flanker performance, while the second patient manifested the opposite pattern, suggesting that their language control impairments (which were similar in both patients) were dissociated from their general executive performance. Dash and Kar (2014) examined the performance of aphasic patients on linguistic and non-linguistic versions of the Flanker task and found that some patients relied more on proactive control in the linguistic task and reactive control in the non-linguistic task, while other patients showed the opposite pattern. Relatedly, Gray and Kiran (2016) examined the performance of aphasic bilingual adults in a word-pair relatedness task with semantically related and unrelated primes, and their performance in the Flanker task. They found dissociation between congruence effects in the two tasks, which they took to indicate that lexical control is domain-specific.

### 1.1.2. Studies addressing the Training Hypothesis

Studies on TH have traditionally relied on a between-groups approach, where the executive performance of bilinguals is pitted against that of monolinguals. Although several individual studies have found better executive performance among bilinguals (for a review, see Bialystok, 2009), recent comprehensive meta-analyses indicate that the bilingual advantage is either negligible or non-existent (e.g. de Bruin et al., 2015; Lehtonen et al., 2017). For example, the meta-analysis by Lehtonen et al. (2017) included 150 studies with 889 effect sizes and examined the hypothesized bilingual advantage across different executive tasks and domains, considering various study-, task-, and participant related variables as moderators. After correcting for bias (publication bias or file drawer effect), no bilingual advantage was found in any executive domain (inhibitory control, shifting, monitoring, working memory, or attention; see also de Bruin et al., 2015; Donnelly, 2016; Hilchey and Klein, 2011; but see Adesope, Lavin, Thompson, and Ungerleider, 2010).

A lack of bilingual advantage is compatible with either of the two following combinations: (i) EH holds and TH does not, or (ii) neither EH nor TH holds. We cannot decide which of these two combinations is true simply based on lack of bilingual advantage, but if there were *independent* evidence supporting EH, then the lack of BEA would suggest that particularly TH is false.

TH can also be examined using a within-group approach, that is, by focusing on associations between language switching frequency and EF within bilinguals. This approach has the advantage of not requiring the matching of two natural groups, which is often difficult. Again, the drawback of this approach is that it is correlative and cannot answer questions about causality: it could either be that subjects with better EF switch between languages more, or that switching more trains EF. However, a positive association would nevertheless be compatible with both EH and TH, whereas a lack of association would not. Earlier studies utilizing this within-group approach have yielded somewhat inconsistent results. Prior and Gollan (2011) found a switching but no mixing advantage in a color-shape task for higher-frequency language switchers, whereas Soveri et al. (2011) found a mixing but no switching advantage in a number-letter task, as well as no inhibitory advantage in the Simon and Flanker tasks. Verreyt et al. (2016), in turn, found an inhibitory advantage in the Simon and Flanker tasks for higher-frequency language switchers. Hartanto and Yang (2016) found a switching but no mixing advantage in a color-shape switching task. Finally, Johnson et al. (2015) did not find any associations between switching frequency and EF measured with Simon, Flanker, and color-shape switching tasks. In sum, the results of earlier studies are inconsistent as to

whether they found an advantage for higher-frequency language switchers, and if they did, in which executive domain or task.

### 1.1.3. Summary of earlier research

In sum, functional neuroimaging studies indicate that language switching relies on the same neuroanatomical regions as general EF tasks, but this does not yet prove that they rely on the same *functions* (many functions can be realized in the same regions, at least given the current spatial resolution of fMRI). Testing whether language switching and general EF tasks rely on the same functions requires using behavioral methods. The few studies that have directly examined the associations between a subject's performance in EF- and language switching tasks have supported EH, but some of them suffer from methodological weaknesses. Studies on aphasic patients indicate dissociations between linguistic and general EF tasks, against the prediction of EH: if both tasks relied on the same general mechanisms, then a damage to these mechanisms should result in deficiencies in both types of tasks. Finally, the lack of bilingual executive advantage suggests either that TH is false, or that both EH and TH are false.

## 1.2. The present study

The studies in this thesis addressed both the Engagement and Training Hypotheses using a within-subjects approach. In other words, we did not pit bilinguals and monolinguals against each other, as is typical in studies on BEA, but instead examined the associations between language switching and EF within bilinguals. Studies 1 and 2 examined the associations between self-reported frequency of everyday language switches in early, balanced Finnish-Swedish bilinguals and their executive performance. The types of language switches assessed in these studies were intended, unintended, and contextual switches, and in Study 2 additionally switches between writing and speech. The Engagement Hypothesis predicts that subjects with better EF would also manifest better language switching and mixing performance, as it can be expected that subjects with better EF make more intended switches and less unintended switches. The Training Hypothesis implies that subjects who have received more training in language switching perform better in executive tasks. Studies 3 and 4, in turn, examined the relationship between executive functions and language switching and mixing costs in laboratory tasks in late, unbalanced Finnish-English bilinguals. The Engagement Hypothesis implies that a subject's performance in the general EF and language switching tasks should

be correlated, due to their relying on the same cognitive mechanisms. Let us next examine the studies in further detail.

To estimate everyday language switching frequency, Study 1 utilized the Bilingual Switching Questionnaire (BSWQ; Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman, & Münte, 2012), whereas Study 2 utilized Ecological Momentary Assessment (EMA; e.g. Shiffman, Stone, & Hufford, 2008), which, to our knowledge, has not been adapted to language switching behavior before. The motivation for using EMA was that general retrospective self-reports of language switching behavior could be prone to errors and biases (e.g. Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Subjects may be poor at estimating how they switch between languages generally and may instead rely on heuristics, such as their language use on a specific occasion they can recall (“availability heuristic”), which may not be representative of their language use more generally. In our variant of the EMA approach, the subjects installed an application on their smartphone that reminded them to report their language switching behavior from the last two hours on roughly two-hour intervals for a period of 14 days. The EMA approach is arguably more valid than general, retrospective questionnaires, as it only probes language switches made during a brief period, which the subject can be expected to recall accurately. Moreover, EMA provides multiple measurements of language switching in the subject’s natural environment, which arguably increases its ecological validity compared to other methods.

In Study 2, our goal was twofold: to examine the associations between language switching (assessed with EMA) and EF, and to assess the convergent validity of BSWQ as well as other general retrospective language switching questions similar to those that have been utilized in earlier studies (e.g. “How often are you in a situation in which you switch between languages?”; Verreyt, Woumans, Vandelandotte, Szmalec, & Duyck, 2016; see also Johnson et al., 2015; Prior & Gollan, 2011).

Studies 3 and 4 addressed directly the Engagement Hypothesis by examining associations between executive performance of bilinguals and their language switching and mixing costs in a picture naming (Study 3) and a semantic categorization (Study 4) task, all performed in laboratory. The Engagement Hypothesis predicts that subjects with better EF would also perform better in the language switching tasks. More specifically, we aimed to test the Inhibitory Control model of bilingual lexical access, which implies that the non-target language is inhibited by a domain-general control mechanism in order to produce the target-language (Green, 1998). The model predicts that the stronger language L1 needs to be inhibited more during processing of the weaker language L2 than vice versa. Moreover, we tested the hypothesis that switching between languages relies on



similar executive mechanisms as non-lexical, general task switching tasks (cf. Meuter & Allport, 1999).

The role of general EF can be expected to be smaller in reception than in production tasks, because in reception lexical access is arguably more stimulus-driven than in production (cf. Peeters, Runnqvist, Bertrand, & Grainger, 2014). Moreover, the IC model pertains specifically to language production. However, a study by Pellikka, Helenius, Mäkelä, and Lehtonen (2015) found larger magnetoencephalography (MEG) responses to dominant language L1 stimuli in a weaker language L2 context than to L2 stimuli in an L1 context in a receptive (semantic categorization) task. This could be taken to suggest that, in this receptive task, L1 is endogenously inhibited in an L2 context to facilitate L2 processing, in line with the IC model. In Study 4, we wanted to further test the possibility that inhibition is involved also in reception by examining the associations between receptive language switching and executive control processes.

### 1.3. Aims and hypotheses

Our aim in the four studies was to assess both the Engagement and Training hypotheses. In Studies 1 and 2, TH implies that bilinguals who in their everyday life make more intended switches also perform better in executive tasks, due to more training. Likewise, EH in these studies predicts that subjects with lower EF make more unintended switches, which can be taken to indicate lapses in executive functioning. It can also be expected that subjects with better EF make more intended switches, because they can switch between languages with less effort. Studies 3 and 4, in turn, mainly address the EH, which predicts that subjects with better general EF would also show better language switching and mixing performance. Moreover, following the IC model, we would expect that better inhibitory capacity correlates with smaller switch cost asymmetry.

## 2. Method

### 2.1. Participants

Key participant characteristics are summarized in Table 1. Studies 1 and 2 employed early balanced Swedish-Finnish bilinguals who use both languages actively. Studies 3 and 4 employed late unbalanced Finnish-English bilinguals who had acquired English at school, roughly at the age of 9. Late unbalanced bilinguals were chosen for Studies 3 and 4 to test the hypothesis that inhibitory control is central specifically in unbalanced bilinguals (Green, 1998; Meuter & Allport, 1999).

Table 1  
*Key participant characteristics*

Study	Exp	Participants ( <i>N</i> )	Mean age (SD, range)	L1	L2	Age of L2 acquisition
1	1	66 early balanced bilinguals	20.3 (3.8, [16, 41])	Fin	Swe	2.4 (1.9, [1, 6])
	2	111 early balanced bilinguals	38.8 (14.9, [18, 69])	Fin	Swe	3.0 (1.8, [1, 6])
2	1	30 early balanced bilinguals	25.1 (5.0, [20, 38])	Fin	Swe	1.3 (1.65, [0, 5])
3-4	1	51 late unbalanced bilinguals	28.6 (7.0, [19, 51])	Fin	Eng	9.1 (1.6, [4, 13])

### 2.2. Procedure

#### 2.2.1. The executive tasks

The general executive tasks in studies 1 and 2 were Flanker and Simon (which tap into general inhibition), visuospatial n-back (working memory (WM) updating), and a number-letter task (set shifting). Studies 3 and 4 utilized the same versions of the Simon, Flanker and number-letter tasks as Studies 1 and 2. The n-back task was not utilized in Studies 3 and 4 because WM updating was not expected to correlate with language switching performance. The EF tasks were administered online in Studies 1 and 2 and in laboratory in Studies 3 and 4. Online testing was conducted using the programmable Soile platform developed at Åbo Akademi University, which runs on the user's web browser using Javascript™. Lab testing was performed with Presentation™ (Neurobehavioral Systems). The executive tasks are briefly described below (Study 1 gives detailed descriptions).

The Simon and Flanker tasks consist of congruent and incongruent trials, and on the latter ones the participant is required to suppress information irrelevant to the task. In the Simon task, the participant is instructed to respond to the color of the stimulus, presented randomly either left or right to the fixation cross, with a left or right key press irrespectively of the location of the stimulus. On congruent trials, the correct response key is on the same side as the stimulus, while on incongruent trials it is on the opposite side. In the Flanker task, the participant's task is to judge whether the central arrow in a row of five arrows points to left or right. On congruent trials, all the arrows point in the same direction; on incongruent trials the central arrow points in the opposite direction than the other four. In both tasks, the relevant measure is a cost effect, defined as incongruent trial reaction time (RT) or error rate minus congruent trial RT or error rate. The higher the cost effect, the weaker the subject's inhibitory capacity is.

In visuospatial *n*-back, the participant's task is to judge whether the stimulus appears at the same location as the stimulus *n* trials before. Our variation of the task included 1-back and 2-back trials. The *n*-back effect is defined as the performance difference between 2-back and 1-back trials, with a larger effect indicating worse WM updating ability.

In the number-letter task, the participant is shown number-letter pairs (e.g. "N3") in one of two squares, one on top of the other. If the stimulus appears in the upper box, the subject has to categorize the letter as a vowel or consonant, and if the stimulus appears in the lower box, the subject has to categorize the number as even or odd. One response key is for vowel letter or even number, another for consonant letter or odd number. The task consists of two single task blocks (letter only and number only), and a mixed task block where task switching is called for. The task produces two key measures: switch and mixing cost. The former is defined as the performance difference between switch and repetition trials in the mixed block, whereas the latter is defined as the difference between mixed block repetition trials and single block trials of the same task. The switch cost represents the cognitive cost of switching between tasks. What the mixing cost reflects is less clearly understood, but it can be considered as a measure of global control processes or task competition management (cf. Rubin & Meiran, 2005).

### 2.2.2. Everyday language switching behavior

Studies 1 and 2 examined the relationship between everyday language switching frequency in bilinguals and their performance in executive tasks. To assess switching frequency, Study 1 utilized the Bilingual Switching Questionnaire (BSWQ; Rodriguez-Fornells et al., 2012), which consists of 12 questions and yields estimates

for four factors: the participant's tendency to switch into L1 or L2 (factors BSWQ-L1S and BSWQ-L2S, respectively), and their tendency to make unintended switches (BSWQ-US) or contextual switches (BSWQ-CS) in their speech. The factor BSWQ-L1S consisted of questions such as "When I cannot recall a word in L2, I tend to immediately produce it in L1"; BSWQ-L2S consisted of the same questions as BSWQ-L1S but with the languages inverted; BSWQ-CS included questions such as "There are situations in which I always switch between the two languages"; and BSWQ-US of questions such as "It is difficult for me to control the language switches I introduce during a conversation". The BSWQ does not specify the time period for which language switching behavior is assessed, rather it asks the participant to "answer to what degree the [...] questions are representative of the manner you use to talk or speak in the languages you know" (Rodriguez-Fornells et al., 2012).

In Study 2, we employed Ecological Momentary Assessment (EMA; e.g. Shiffman, Stone, & Hufford, 2008) to measure the participants' language switching behavior. The participants' smartphone reminded them to answer four language switching questions six times a day on roughly two-hour intervals for two weeks. The questions prompted the participants to estimate how many intended (EMA-IS) and unintended (EMA-US) language switches they made during the last two hours in their speech. Additionally, they estimated the typicality of their language switches (EMA-CS), and how many switches they had made between writing and speech (EMA-WS). The questions are summarized in Table 2. After the two-week EMA period, the subjects rated how well their language switching behavior during that period corresponded to their general switching behavior. This question was included to be able to make broader conclusions than just for the two weeks they participated in EMA.

In addition to examining associations between everyday language switching and EF, Study 2 also investigated the validity of the BSWQ and individual language switching questions similar to those that have been used in earlier studies (e.g. Johnson et al., 2015; Prior & Gollan, 2011; Verreyt et al., 2016). The individual language switching questions included in Study 2 were "On average I switch between languages during a day [x times]" (Average Switching Frequency, ASF), and "On average I make many brief language switches during a day" [agree-disagree] (Many Brief Switches, MBS). The subjects responded to BSWQ, ASF, and MBS before and after the EMA period. The pre-post ratings were used to assess the reliability of these general retrospective question(naire)s. To assess validity, the pre-EMA ratings in the BSWQ factors, ASF, and MBS were compared to the EMA questions that most closely resembled them. The protocol of Study 2 is summarized in Figure 2.

Table 2

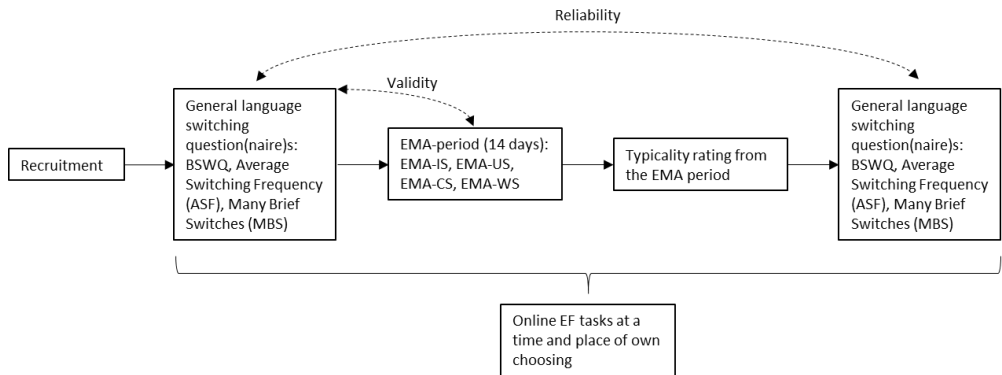
*The questions used in the Ecological Momentary Assessment period and their codings<sup>a, b</sup>*

Question	Abbreviation	Question	Scale	Re-coding
EMA-Intended Switches	EMA-IS	“After your last answer (or during the last max two hours), how many times did you intentionally switch between languages in your speech?”	0, 1-3, 4-6, 7 or more often	0 = 1; 1-3 = 2; 4-6 = 3; 7 or more often = 4
EMA-Unintended Switches	EMA-US	“After your last answer (or during the last max two hours), how many times did you unintentionally switch between languages in your speech?”	0, 1-3, 4-6, 7 or more often	0 = 1; 1-3 = 2; 4-6 = 3; 7 or more often = 4
EMA-Contextual Switches	EMA-CS	“How large proportion of the aforementioned language switches took place in situations in which you typically always switch between languages?”	0-25%, 25-50%, 50-75%, 75-100%	0-25% = .125; 25-50% = .375; 50-75% = .625; 70-100% = .85
EMA-Switches Between Writing and Speech	EMA-WS	“After your last answer (or during the last max two hours), how many times did you switch between languages in your writing, or between writing and speech?”	0, 1-3, 4-6, 7 or more often	0 = 1; 1-3 = 2; 4-6 = 3; 7 or more often = 4

<sup>a</sup> The re-coded values were used in the analysis, except for the EMA-CS variable which was further recoded for analysis as follows: Estimated frequency of contextual switches = EMA-CS as coefficient  $\times$  (EMA-IS + EMA-US). In what follows, I will denote the final variable simply as EMA-CS.

<sup>b</sup> The specification “or during the last max two hours” in the questions was added because the period from the previous response was sometimes longer than two hours, e.g. in the first response of the morning when the last response was from the previous evening. We did not want the subjects to estimate their switching from periods longer than two hours, because we assumed that this would have lessened the accuracy of the answers.

Figure 2. Setup in Study 2.



### 2.2.3. Language switching laboratory tasks

Studies 3 and 4 examined whether the executive performance of late unbalanced Finnish-English bilinguals is related to their language switching performance in laboratory. In Study 3, the participants named photos of everyday objects in Finnish or English, depending on the picture background color. The picture naming task included Finnish and English single language blocks where the pictures were named only in Finnish or English, and a mixed language block where the naming language was switched pseudorandomly on intervals of two to four same-language trials. The picture naming task yields measures of the participant's language switching and mixing capacity. The switch cost is defined as performance difference between switch and repetition trials in the mixed language block. The mixing cost is defined as the performance difference between single block trials in a language L and mixed block repetition trials in L. The mixing cost can be considered as a monitoring or alertness cost: even though in the mixed block repetition trials the language does not shift, the subject is prepared that it could shift, which can be expected to cause longer reaction times or higher error rates in the repetition trials than in the single block trials. Higher switch and mixing costs indicate worse language switching or mixing capacity. Study 4 was identical in setup to Study 3, with the exception that the language switching task was a semantic categorization task (i.e., a reception task) where participants read words one at a time on a screen and decided whether the object the word denotes is animate or inanimate.

The next section summarizes the results of the four studies. For details, the reader is advised to consult the original articles.

## 3. Results

### 3.1. Everyday language switching and executive performance

We examined the associations between everyday language switching behavior and EF in multiple regression models with one of the EF cost effects (Simon, Flanker, or n-back effect, or number-letter switch or mixing cost) in either RT or error rate per model as a dependent variable and the language switching measures as independent variables. Thus, 10 models were run in each of the three experiments (two experiments in Study 1 and one in Study 2), resulting in altogether 30 models. In both experiments included in Study 1, the predictors were BSWQ-Language Switches (BSWQ-LS; sum score of switches into L1 and switches into L2), BSWQ-Unintended Switches (BSWQ-US), and BSWQ-Contextual Switches (BSWQ-CS). In Study 2 the predictors were the EMA variables EMA-Intended Switches (EMA-IS), EMA-Unintended Switches (EMA-US), EMA-Contextual Switches (EMA-CS), and EMA-switches between Writing and Speech (EMA-WS). The results of Studies 1 and 2 are summarized in Table 2.

In Experiment 1 of Study 1, none of the associations were significant. In Experiment 2, higher BSWQ-US predicted larger n-back effect in error rates, and higher BSWQ-CS predicted larger number-letter switch cost in RT. In Study 2, higher EMA-US predicted larger n-back effect in RT. In sum, most of the models (27 out of 30) were not significant, and in the models that were significant (3 out of 30), US and CS were associated with worse executive performance (in WM updating and set shifting).

#### 3.1.1. Reliability and validity of General Language Switching Questions

In Study 2, we additionally assessed the test-retest reliability and validity of the general language switching question(naire)s BSWQ, ASF (“On average I switch between languages during a day”), and MBS (“On average I make many brief language switches during a day”). In terms of Pearson test-retest correlations, BSWQ-US showed high ( $r = .80, p < .001$ ), BSWQ-LS marginal ( $r = .62, p < .001$ ), and BSWQ-CS non-significant ( $r = .27, p > .1$ ) reliability, whereas the reliabilities were low for ASF ( $r = .40, p < .05$ ) and MBS ( $r = .53, p < .01$ ) (cf. Strauss, Sherman, & Spreen, 2006). To assess the convergent validity of the BSWQ factors, BSWQ-LS was compared to EMA-IS, BSWQ-US to EMA-US, and BSWQ-CS to EMA-CS. The simple questions ASF and MBS were compared to EMA-IS. Of the BSWQ factors, only BSWQ-US showed high correlation with EMA-US ( $r = .63, p < .001$ ), while convergent validities for BSWQ-LS ( $r = -.14, p > .1$ ) and BSWQ-CS ( $r = -.15, p > .1$ )

were low and not significant. Of the individual switching questions, ASF correlated moderately with EMA-IS ( $r = .42, p < .05$ ), while MBS did not ( $r = .014, p > .1$ )



Table 2  
*Associations between everyday language switching and EF in Studies 1 and 2. Estimates for individual predictors in a model are reported only if overall model fit (F) was significant.<sup>a</sup>*

		Simon effect				Flanker effect				N-back effect				Switch cost				Number-letter				
		RT		ER		RT		ER		RT		ER		RT		ER		RT		ER		
		RT	ER	RT	ER	RT	ER	RT	ER	RT	ER	RT	ER	RT	ER	RT	ER	RT	ER	RT	ER	
Study 1, Experiment 1	Model fit	F	2.33	.23	.24	.070	.20	1.64														
	indices	R <sup>2</sup> <sub>a</sub>	.059	-.038	-.038	-.046	-.066	.048														
	BSWQ	LS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	predictors (B [95%CI])	US	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Study 1, Experiment 2	Model fit	F	.14	.56	.12	1.06	1.29	2.83*														
	indices	R <sup>2</sup> <sub>a</sub>	-.025	-.013	-.026	.002	.009	.047														
	BSWQ	LS	-	-	-	-	-	-.74 [-2.26, .79]														
	predictors (B [95%CI])	US	-	-	-	-	-	2.61* [.61, 4.61]														
Study 2	Model fit	F	1.03	1.44	1.22	.52	3.62*	2.18														
	indices	R <sup>2</sup> <sub>a</sub>	.0040	.058	.030	-.071	.29	.15														
	EMA	IS	-	-	-	-	-	-70.03 [-331.83, 191.78]														
	predictors (B [95%CI])	US	-	-	-	-	-	141.09* [7.99, 274.19]														
		CS	-	-	-	-	-	-96.45 [-234.23, 41.33]														
		WS	-	-	-	-	-	-58.16 [-197.07, 80.73]														

<sup>a</sup> ER = Error Rate. Log-transformed RTs were used in Study 1 to meet the normality assumption; raw RTs were used in Study 2. Asterisk indicates statistical significance (\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ ).

### 3.2. Language switching in laboratory and executive performance

Studies 3 and 4 examined the associations between the language switching and mixing performance of late unbalanced bilinguals and their executive performance. Study 3 utilized a picture naming task and Study 4 semantic categorization. In both studies, language switch and mixing costs were analyzed using linear mixed effects models with log-transformed RT as dependent variable and Condition (mixed-block repetition, mixed-block switch, or single block repetition) and Language (Finnish or English) as predictors. Associations between language switch and mixing costs and EF cost effects were analyzed in models that included, in addition to the aforementioned variables, one of the EF cost effects in log-transformed RTs as predictor per model. Table 3 summarizes the results in Studies 3 and 4. Findings in line with the hypotheses are marked with “+”, those in conflict with the hypotheses are marked with “-”.

In both production and reception, we expected switch trials to be responded to more slowly than repetition trials in both languages (i.e., a positive switch cost); we also expected mixed block repetition trials to be responded to more slowly than single block trials in both languages (i.e., a positive mixing cost). Following the Inhibitory Control model (Green, 1998), we hypothesized that switch and mixing costs would be larger for L1 than for L2, because the model implies that L1 is inhibited more strongly than L2 in the mixed block. This is called switch or mixing cost asymmetry (Green, 1998; Meuter & Allport, 1999). As to the interactions with the EF tasks, we expected better inhibitory capacity (smaller Simon or Flanker effect) to predict better language switching capacity (smaller language switch and mixing costs), in line with the IC model. As to the number-letter task, we focused on the relationship between the number-letter switch effect (NLSE) and the language switch cost, and the number-letter mixing effect (NLME) and the language mixing cost (i.e., we did not examine the relationship between e.g. NLME and language switch cost, because we assumed mixing and switching to be different types of processes). We expected better set-shifting capacity to be associated with better language switching and mixing performance (cf. Meuter & Allport, 1999).

From Table 3 we can see that in the picture naming task, all the basic language switch and mixing effects, plus asymmetry, were in line with the hypotheses. The significant interactions of the switch and mixing costs with the Simon effect were, however, against the hypotheses: Simon correlated negatively with the L1 mixing cost and L2 switch cost. The number-letter switching effect, however, positively predicted the L2 switch cost.

In the semantic categorization task, there was a positive switch cost in both languages of symmetric magnitude, but no mixing cost in L1, and a strong mixing advantage in L2. That is, contrary to our hypotheses, L2 mixed block repetition trials were responded to faster than L2 single block trials. As to the interactions between the executive cost effects and language switch and mixing costs, in L1 higher Flanker effect predicted larger mixing cost. In L2, however, both Flanker and Simon predicted smaller (more negative) mixing cost, that is, larger mixing advantage.

Finally, the number-letter mixing effect predicted larger (more positive) mixing cost in both languages.

Table 3

*Associations between language switching in picture naming (production) and semantic categorization (reception), and the EF cost effects in Studies 3 and 4<sup>a,b</sup>*

	Production			Reception		
	E	SE	t	E	SE	t
<b>Language switch and mixing costs</b>						
<b>L1</b>						
Switch cost	<b>+0.023</b>	<b>.0020</b>	<b>11.66***</b>	<b>+0.012</b>	<b>.0030</b>	<b>3.95***</b>
Mixing cost	<b>+0.036</b>	<b>.0014</b>	<b>25.48***</b>	-.0029	.0018	1.58
Switch cost × L2 (asymmetry)	<b>+0.0062</b>	<b>.0028</b>	<b>2.20*</b>	+0.0039	.0042	.92
Mixing cost × L2 (asymmetry)	<b>+0.036</b>	<b>.0020</b>	<b>17.67***</b>	<b>+0.020<sup>c</sup></b>	<b>.0026</b>	<b>7.74***</b>
<b>L2</b>						
Switch cost	<b>+0.017</b>	<b>.0020</b>	<b>8.45***</b>	<b>+0.0081</b>	<b>.0030</b>	<b>2.70**</b>
Mixing cost	+0.00011	.0015	.075	<b>-.023</b>	<b>.0018</b>	<b>12.51***</b>
<b>Interactions between the language switch and mixing costs and the EF cost effects</b>						
<b>L1</b>						
Switch cost × Simon	+0.055	.098	.56	+0.016	.13	.13
Mixing cost × Simon	<b>-.33</b>	<b>.071</b>	<b>4.65***</b>	-.016	.095	.17
Switch cost × Flanker	-.18	.11	1.70	+0.12	.13	.98
Mixing cost × Flanker	+0.0026	.078	.033	<b>+0.47</b>	<b>.094</b>	<b>5.03***</b>
Switch cost × NLSE	+0.049	.040	1.21	-.041	.047	.87
Mixing cost × NLME	-.0041	.024	.17	<b>+0.16</b>	<b>.031</b>	<b>5.13***</b>
<b>L2</b>						
Switch cost × Simon	<b>-.24</b>	<b>.099</b>	<b>2.41*</b>	+0.070	.13	.55
Mixing cost × Simon	-.12	.073	1.62	<b>-.27</b>	<b>.095</b>	<b>2.85**</b>
Switch cost × Flanker	-.15	.11	1.38	-.13	.13	1.06
Mixing cost × Flanker	-.093	.081	1.15	<b>-.35</b>	<b>.094</b>	<b>3.71***</b>
Switch cost × NLSE	<b>+0.083</b>	<b>.041</b>	<b>2.01*</b>	-.0055	.046	.12
Mixing cost × NLME	-.0081	.025	.74	<b>+0.10</b>	<b>.031</b>	<b>3.37***</b>

<sup>a</sup> Significant effects are boldfaced. All effects in line with the hypotheses are marked with “+”, all effects that contradict the hypotheses are marked with “-”. One model per domain (production or reception) was used to analyze the language switch and mixing costs. The interactions with EF cost effects were obtained from multiple models: one model per EF cost effect. The table, thus, summarizes altogether 10 models.

<sup>b</sup> NLSE = number-letter switching effect; NLME = number-letter mixing effect.

<sup>c</sup> The L2 mixing cost asymmetry is coined as being according to hypotheses (“+”), but it is only trivially so: the mixing cost was negative in both languages, but it was less negative (i.e., more positive) in L1.

## 4. Discussion of the individual studies

We tested the hypotheses that bilingual language switching engages and trains executive functions in four studies. In studies 1 and 2, we examined the associations between everyday language switching in early balanced bilinguals and their executive performance. The Training Hypothesis implies that a higher rate of (intentional) language switching would be associated with higher executive performance (due to more practice). The Engagement Hypothesis, in turn, can be taken to imply both that subjects with higher EF make more intended language switches, as well as that poorer EF would be associated with more unintended language switching. In Studies 3 and 4, we examined the associations between executive functioning, and language switching and mixing performance of late unbalanced bilinguals in two laboratory tasks (picture naming and semantic categorization). The Engagement Hypothesis implies that, if language switching utilizes general EF, subjects with better EF (smaller EF cost effects) would show better language switching and mixing performance (smaller language switch and mixing costs).

### 4.1. Everyday language switching and executive performance

Studies 1 and 2 provide modest evidence for an association between unintended language switching in everyday life and poorer WM updating capacity, in line with the Engagement Hypothesis. This can be taken to indicate that subjects who have poorer WM monitoring and updating capacity are also worse at monitoring the activation levels of their languages and thus make more unintended language switches. In turn, the association between BSWQ-CS and worse WM updating performance in Study 1 could be a false positive finding, due to the lack of validity of the BSWQ-CS factor. The results do not support the Training Hypothesis, which implies that subjects who (intentionally) switch between languages more in their everyday life perform higher on the EF tasks (due to more practice).

The results on the validity of BSWQ and the individual language switching questions in Study 2 bring into doubt the results of earlier research that have utilized such methods. Only the BSWQ-US factor showed high convergent validity; other factors showed modest or non-significant validity. As to studies that have used BSWQ, some have found significant associations between higher rate of language switches (BSWQ-LS) and better EF (Rodriguez-Fornells et al., 2012; Soveri et al., 2011). Given the lack of validity of BSWQ-LS, these findings might be false positives and cannot be taken as evidence for the Training Hypothesis. Besides BSWQ, previous studies have utilized individual language switching questions similar to Average Switching Frequency, which had only moderate validity (Johnson et al., 2015; Prior & Gollan, 2011; Verreyt et al., 2016). The earlier studies have yielded inconsistent associations between EF and switching frequency: some studies have found a significant association while others have not, and the positive findings have been found in different types of EF tasks. These inconsistencies could be due to the

relatively low validity of the language switching questions used in the earlier studies, which motivates the introduction of more valid methods such as EMA.

From a more general perspective, the results of Studies 1 and 2 indicate that language switching does not train general EF, in line with the results of recent meta-analyses indicating no bilingual executive advantage (de Bruin et al., 2015; Lehtonen et al., 2017). It is important to keep in mind here that the approach of these studies was correlational and cannot answer questions about causality (whether more language switching leads to better EF or vice versa). For instance, higher EF could lead to more intended language switching because it makes switching less effortful. As to the Engagement Hypothesis, the implications of Studies 1 and 2 are clearer: it is more probable that low WM updating capacity leads to more unintended switches than that making more unintended language switches leads to lower WM updating capacity (on the assumption that one's language switching behavior cannot worsen their WM capacity). Overall, the results indicate that language switching may be linked to WM updating (in line with EH) but does not suggest that higher switching frequency would train WM updating (in contrast to TH). These results should be interpreted with caution, however, because no previous research (to my knowledge) has examined the associations between switching frequency and WM updating. Although the effect size in Study 2 was relatively high ( $R^2_a = .29$ ), the sample size was quite small ( $N = 30$ ), prompting the replication of the study in a larger sample.

## 4.2. Language switching in laboratory and executive performance

In the picture naming task, we found significant language switch and mixing costs in the expected direction. Moreover, the switch cost was larger for L1 than L2, an asymmetry typically taken as evidence of inhibition of L1 (cf. Bobb & Wodniecka, 2013). However, the interactions of the language switch and mixing costs with the EF cost effects showed inconsistent results. The IC model implies that the stronger L1 is suppressed more during the production of the weaker L2 than vice versa. Hypothetically, this would lead to a larger cost when switching into L1 than when switching into L2, because switching into L1 from L2 requires resolving the inhibition of L1 from the previous L2 trial (Green, 1998). Thus, we expected the inhibitory cost effects to correlate particularly with the L1 switch cost. In an earlier study utilizing a similar approach, Linck et al. (2012) found a positive association between L1 switch costs and the Simon effect.

Contrary to our hypotheses and the results of the Linck et al. (2012) study, neither of the inhibitory control tasks (Simon or Flanker) predicted the L1 switch cost. Larger Simon effect did, however, predict smaller L1 mixing cost. The opposite would have been expected if the mixing cost was due to inhibition of L1 in the mixed block. Moreover, larger Simon effect predicted smaller switch cost in L2. This finding was surprising in two ways: first, inhibition is not considered as central when switching into L2; second, even if it was, we would have expected an effect in the opposite direction. The only EF interaction in line with our hypotheses was a

significant association between larger number-letter switching effect and L2 language switch cost, which can be taken to indicate that better task switchers are better at switching into L2.

All in all, although the basic language switching and mixing costs were in line with earlier research and with the IC model, the EF interactions either did not support or were against the IC model and the Engagement Hypothesis. The results lend some (weak) support for the hypothesis that language switching is linked to general set shifting, but this evidence was limited to switching into L2. It could be speculated that whereas switching into the dominant L1 is an automatized process, switching into L2 is less automatic and engages the subject's general set shifting capacity more. However, if switching into L1 was more automatized than switching into L2, then assumedly the L2 switch cost should be larger than the L1 switch cost, which is typically not observed (Bobb & Wodniecka, 2013). Moreover, if switching into L2 was particularly loading on general set shifting, then we would expect that it would also positively correlate with the subject's inhibitory capacity (because set-shifting arguably also engages inhibition, i.e., inhibiting the non-target task set). We found, in contrast, that the Simon effect correlated *negatively* with the L2 switch cost. Thus, the EF interactions in the production task can be considered as inconsistent and do not support the Engagement Hypothesis.

In the semantic categorization task, we found a significant switch cost in both languages in the expected direction. The switch costs were of symmetrical strength (cf. Macizo, Bajo, & Paolieri, 2012; von Studnitz & Green, 2002; but see Jackson, Swainson, Mullin, Cunnington, & Jackson, 2004; Pellikka et al., 2015). The IC model implies that if inhibition was engaged in receptive language switching, this would lead to larger switch costs for L1 than for L2. Thus, the lack of asymmetry can be taken to be in line with the hypothesis that lexical access in the semantic categorization task is not driven by executive control processes (cf. Dijkstra & van Heuven, 2002; van Heuven & Dijkstra, 2010).

There was no mixing cost in L1, but, unexpectedly, a large mixing advantage in L2: mixed block repetition trials were responded to faster than single block trials. The L2 mixing advantage can be considered as evidence of statistical facilitation, in line with a horse-race model (Raab, 1962). The response (animate/inanimate) in the semantic categorization task does not depend on the stimulus language. Thus, it can be speculated that in the mixed block the subjects can use the lexical routes of both L1 and L2 in parallel to make a response, and the response is always made through the route that first reaches the "goal". This leads to facilitation when the reaction time distribution for L1 and L2 overlap (for a more detailed description, see Study 4). The facilitation occurs specifically in the mixed block where the lexical routes of both languages are active, not in the single language block where only one language is active. Another approach to interpret the L2 mixing advantage is in terms of the Adaptive Control hypothesis (Green & Abutalebi, 2013), which implies that a bilingual may use the lexical route of either language to make a response in a mixed language context, a process Green and Abutalebi (2013) title "opportunistic planning". We discuss these possibilities in more detail in the article of Study 4.

From a more general perspective, these accounts of the L2 mixing advantage presuppose that lexical access is non-selective—in other words, they imply that the non-target language is not selectively inhibited. Thus, the L2 mixing advantage can be taken as evidence that top-down control processes do not play a role in language reception, but instead that lexical access in reception is non-selective.

The interactions between the language switch and mixing costs in the semantic categorization task and the executive cost effects were, again, inconsistent. In Table 3, the interactions are marked with “+” if the language switching or mixing cost was larger (more positive) the larger the EF cost effect. Because there was no L2 mixing cost but instead a mixing advantage, an interaction marked with “+” indicates that the mixing advantage becomes smaller (i.e., the switch cost becomes more positive) when the EF cost effect increases. In L1, the Flanker effect predicted larger (more positive) mixing cost, which could be taken to indicate that L1 is selectively suppressed in the mixed block during the reception of L2. However, such suppression would arguably also lead to a positive interaction of the Simon task with the L1 switch cost, which was not observed. Moreover, the inhibitory model conflicts with the L2 mixing advantage, if that is taken as evidence of non-selective lexical access. In L2, both the inhibitory tasks predicted a smaller (less positive) switch cost, that is, the L2 mixing advantage was higher the weaker the subject’s inhibitory capacity. It could be argued that subjects with weaker inhibitory capacity relied less on inhibition of the target language and more on between-language facilitation (or opportunistic planning; Green & Abutalebi, 2013) to make a response. However, this suggestion is not compatible with the finding that in L1 mixing cost was larger (more positive) the higher the Flanker effect. The only consistent EF interaction in Study 4 was that between the number-letter mixing effect and the language mixing cost: in both languages, higher number-letter mixing effect was associated with larger (more positive) language mixing cost. This could be taken to indicate that subjects who were better in general task mixing were also better language mixers.

All in all, in Study 4 the interactions of the language mixing costs with the inhibitory cost effects were inconsistent. Interpretation of the results is even more complex because an L2 mixing advantage instead of a cost was found, and we had no predictions as to how a mixing advantage would be associated with general EF. The only consistent interactions were between the number-letter and language mixing costs, indicating that better task mixers are also better language mixers. However, even the latter finding is to be interpreted with caution, because in L2 there was no mixing cost, but instead a mixing advantage. That is, in L2 it was not the case that the mixing cost was smaller the smaller the number-letter mixing cost; instead, the mixing advantage was larger the smaller the number-letter mixing cost.

The results of studies 3 and 4 do not support the Engagement Hypothesis and contradict the results of earlier research (Linck et al., 2012; Liu et al., 2016, 2014). With respect to the Liu et al. studies, our approach arguably had better power to detect significant findings, due to treating EF performance as a continuous variable. The Liu et al. studies dichotomized the subjects based on their EF performance, and only examined the presence of language switch cost asymmetry in both groups, not

the direct (assumedly linear) associations between language switch costs and executive performance. The Linck et al. (2012) study, in turn, utilized only ten black-and-white drawings as stimuli, which were repeated 60 times each during the picture naming task. It can be argued that this setup is particularly loading on inhibition (see section 1.1.). Finally, we utilized three different executive tasks in contrast to just one in the previous studies, which enables to better see if any significant associations are consistent.

The results of Studies 3 and 4 are also somewhat in contrast to the brain imaging results, which indicate that language switching and general EF tasks rely on the same brain regions. However, the behavioral approach arguably better answers the question whether language switching and general EF tasks rely on the same cognitive functions, as it is probable that many different functions can be realized in roughly the same anatomical regions.

### 4.3. Inhibition and asymmetry

As a general remark on Studies 3 and 4, one can question the common assumption that larger L1 switch costs are evidence of inhibition (Bobb & Wodniecka, 2013). Switch cost asymmetry is typically derived from the IC model as follows: the stronger L1 needs to be inhibited to process the weaker L2, and when switching from L2 to L1, the L2 inhibition from the previous trial needs to be resolved. However, the L2 inhibition also needs to be *engaged* when switching from L1 to L2. This raises the question of why resolving inhibition should take more time than the engagement of inhibition. If both engaging and resolving inhibition required an equal amount of processing, then switch costs into L1 and L2 should be symmetrical. One could argue that whereas the engagement of inhibition is an active and fast process, the resolution is a passive process, where the inhibition fades out when not actively sustained. This would arguably lead to an asymmetry in line with the IC model: the cost would be larger when switching into L1 (when the L1 inhibition from the previous L2 trials is slowly fading) in contrast to when switching into L2 (when L1 inhibition is quickly activated). According to this logic, however, a subject's general inhibitory capacity should correlate mainly with the engagement of inhibition (which is an active process), not with its resolution (which is passive). Thus, on this account, subjects with a better inhibitory capacity should manifest smaller costs specifically when switching into L2 (when L1 inhibition is actively engaged), not when switching into L1 (when L1 inhibition from previous L2 trials is passively fading). Our results did not support this alternative view, but this possibility may be worth considering in future studies.



## 5. General discussion

The Bilingual Executive Advantage hypothesis implies that bilinguals outperform monolinguals on executive functions. A possible mechanism underlying this advantage is bilinguals' language switching and mixing, which has been proposed to engage general EF (e.g. Rodriguez-Fornells et al., 2006). In the present thesis, I have investigated whether language switching and mixing engage or train EF. I have argued that the Engagement Hypothesis is a necessary but not sufficient condition for training to occur: simply loading a cognitive mechanism might not improve its performance.

Overall, the results lend only weak support for the Engagement Hypothesis, and no support for the Training Hypothesis. The findings are consistent with recent meta-analyses and systematic reviews that have failed to find consistent evidence of a bilingual advantage. The first meta-analysis on the bilingual advantage by Adesope, Lavin, and Thompson, (2010) found a small-to-large bilingual advantage in several cognitive domains, including attentional control and WM. Hilchey and Klein (2011), in their systematic review of 13 articles, in turn, conclude that advantages are mainly found in older adults. Paap, Johnson, and Sawi (2015) examined the studies published after the Hilchey and Klein (2011) review and found that a bilingual advantage in inhibition or set-shifting was only found in a small portion of the studies (13-22% of the studies). The meta-analysis by de Bruin et al. (2015) found a small bilingual executive advantage ( $d = .30$ ; distinct types of EF tasks pooled), and evidence of a publication bias. A meta-analysis by Donnelly (2016) found a small bilingual advantage in inhibitory control but no evidence of an advantage in set-shifting. Given the indications of a publication bias (de Bruin et al., 2015), true effect sizes are probably even smaller. Finally, the most comprehensive meta-analysis to date by Lehtonen et al. (2017), distinguishing between a range of executive tasks and correcting for publication bias, did not find an executive advantage in any of the executive domains they examined.

Importantly, the lack of a bilingual executive advantage is compatible with the Engagement Hypothesis: language switching might engage but not train general EF. While in the light of present evidence it appears that TH is false, there is evidence both for and against EH, summarized in Table 4 (see also section 1.1.). Acknowledging that the truth of EH is an open question, let us suppose for a moment that EH were false; what alternatives do we have?

Table 4  
*Evidence for and against the Engagement Hypothesis*

	Evidence	Problems
For EH	Brain imaging studies	Low spatial resolution, does not directly answer questions about functionality
	Behavioral studies (Linck et al., 2012; Liu et al. 2014, 2016)	Experimental setup (Linck et al.), low-powered analyses (Liu et al.)
	Study 2 in the present thesis	Positive effects only in WM updating, which is not typically considered central in language switching (unlike e.g. inhibition)
Against EH	Lack of BEA	Could be due to falsity of TH, not EH
	Dissociations between linguistic and non-linguistic EF-demanding tasks in aphasic patients	Partially inconsistent evidence
	Studies 3 and 4 in the present thesis	Inconsistent results could be due to problems in convergent validity between the EF tasks

## 5.1. Within-lexicon control processes

In language reception, theoretical models of lexical access have typically relied either on the assumption of non-selective lexical access (the BIA+ model; van Heuven & Dijkstra, 2010) or within-lexicon control processes, such as inhibitory or excitatory connections between lexical nodes (the BIA model; Dijkstra & van Heuven, 2002). Both models imply that general executive processes are not utilized in language reception. Whereas the BIA models may be plausible in the case of language reception, they arguably cannot be applied to language production. Both models imply that either external stimuli (the BIA+ model) or previously activated lexical representations (the BIA model) determine which lexical representations are consequently activated, which appears to be incompatible with the fact that a person can decide to, say, switch into English in a Finnish conversation. Thus, it appears that some extra-lexical control processes are at play in language production. I will next examine what such processes could be, supposing that they are not general executive processes.

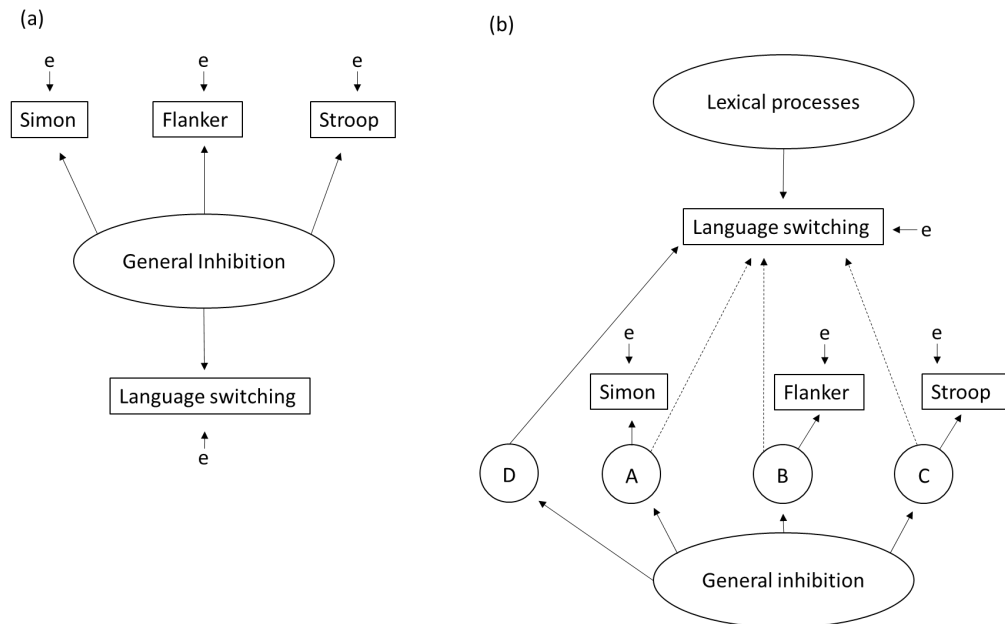
## 5.2. Task-specific control processes

There is evidence that the commonly used executive tasks lack convergent validity (cf. Paap & Sawi, 2014). This is often called the “task impurity problem”, and is typically considered to indicate problems in the validity of executive tasks. In other

words, executive tasks do not measure only what they are intended to, but additionally a range of other phenomena (Burgess, 1997). Instead of impurity, however, this phenomenon could be taken to indicate an essential property of the executive system, namely that processes underlying executive tasks are more task-specific than is typically considered, and load less on a single, general executive faculty.

Consider the influential paper by Miyake et al. (2000), which is often used as a basis for the threefold classification of executive functions into set-shifting, inhibition, and WM monitoring and updating. The authors conducted latent factor analysis of six different executive tasks: Plus-Minus, Number-Letter, and Local-Global (typically taken to tap into set-shifting); Keep Track, Tone Monitoring, and Letter Memory (typically considered as WM updating/monitoring tasks); and Antisaccade, Stop-Signal, and Stroop (assumed measures of inhibition). Confirmatory factor analysis with three latent factors Shifting, Updating, and Inhibition indicated that the tasks loaded on the factors as expected. However, the factor loadings ranged between .33 and .63, which are throughout lower than the error terms in the model, which ranged from .78 to .94. That is, the factor loadings were modest, and most of the variance in the executive tasks was unexplained by the model. Typically, this is considered as noise or error variance which is standardly present in any cognitive tasks. However, the lack of explanatory power of the latent factor solution could also indicate that the different executive tasks rely only partly on a general executive function, and largely on task-specific processes.

Figure 3. Model (a): The traditional view of how a general executive process (in this case general inhibition) is related to specific executive tasks and language switching; model (b): an alternative view, relying on task-specific subsystems. Error terms are denoted by “e”.



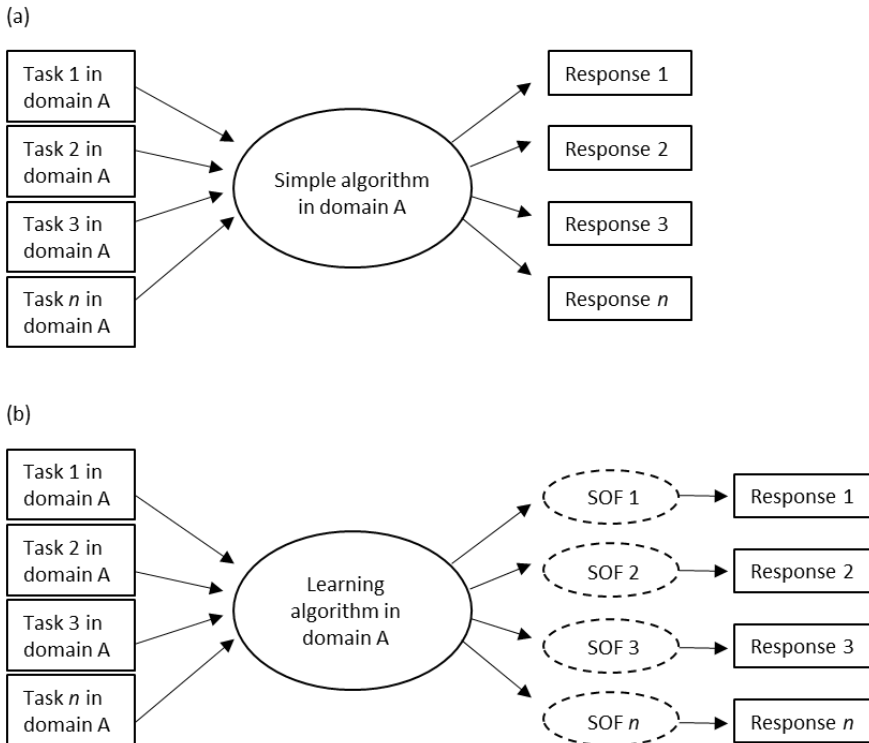
To illustrate, let us examine the two models in Figure 3. Model (a) represents the traditional view (cf. Miyake et al., 2000), where general inhibition directly determines a subject's performance in particular inhibitory tasks. Task impurity stems from error variance ("e" in Figure 3) that is specific to each task, but in principle, on this account, we could devise better inhibitory tasks that more directly tap into general inhibition. As to language switching, model (a) implies that general inhibition is employed. In model (b), in contrast, each inhibitory task has its specific control mechanism (A, B, or C), all of which have some variance in common due to their links to general inhibition. The crucial difference between the two models is that (a) implies that a lack of correlation between different inhibitory tasks is due to error variance, whereas in (b) it is due to the actual layout of the executive system, not just error variance.

Let us focus more closely on how the two models would explain the relationship between language switching and EF. Model (a) implies that the type of executive process employed in language switching is general inhibition, whereas model (b) presumes that the executive component is task-specific (D). It is possible that this subsystem is even more specialized than the subsystems A, B, and C underlying the inhibitory tasks, because language switching is highly automated through vast experience. Thus, D might correlate with A, B, and C to a lesser extent than A, B, and C do with each other. Moreover, language switching is arguably affected by non-executive systems, such as lexical mechanisms, which further lessens the common variance between language switching and the inhibitory tasks. The subsystems A, B, and C might be involved to some slight degree in language switching (the dashed lines in Figure 3) but they need not be; this depends on how structurally similar the language switching and inhibitory tasks are. The main source of common variance between language switching and the inhibitory tasks is through general inhibition, which underlies each of the subsystems A through D. However, the associations between language switching and specific inhibitory tasks are much lower in model (b) than in model (a), due to the presence of subsystems in model (b), which account for a large (or even most) part of the variation in the tasks.

We do not know in which way the task-specific subsystems would be realized, but we can speculate that they would be learnt. An executive process such as "inhibition" can be considered as a function which, given a certain input, produces an output. In information processing terms, a function is realized by a specific algorithm. We can distinguish between (at least) two types of algorithms, simple and learning ones. A simple algorithm is simply a set of rules that determines output as a function of its input, and this same set of rules is utilized in any task where the algorithm is utilized. A learning algorithm, in contrast, produces a task-specific, second-order function to produce an output (cf. deep learning neural networks in computer science). For example, let us suppose that different kinds of inhibitory tasks (Simon, Flanker, Stroop etc.) rely on the same "inhibition algorithm" due to their structural similarities. A simple algorithm uses the same set of rules to produce an output

across all tasks, whereas a learning algorithm produces a unique solution to each task. In other words, the learning algorithm produces the optimal function for responding to the Simon task, learns another optimization for the Flanker task, etc. In both the case of simple and learning algorithms, we may presume that a subject's performance in different tasks within a domain correlate to some extent (due to their relying on the same algorithm), but in the case of learning algorithms the inter-task correlations can be presumed to be lower, due to the unique solution that is utilized in each task. The difference between simple and learning algorithms is illustrated in Figure 4.

Figure 4. Simple (a) vs. learning algorithm (b). The input of the algorithm (on left) are different tasks within a specific domain A (e.g. Simon, Flanker, Stroop, etc. in inhibition). The circle with continuous outline represents the algorithm (function) utilized in all tasks in that domain. The simple algorithm produces a response directly as a function of the input, whereas the learning algorithm produces a unique second-order function ("SOF"; circles with dashed outlines) to each problem, which eventually determines the response.



Both in the case of simple and learning algorithms a subject's performance in different tasks within the same domain would correlate. In the case of simple algorithms this is due to the same function utilized in each task; in the case of learning algorithms, this is due to the same learning mechanism utilized to produce

a second-order function to each task in that domain. However, the inter-task correlations can be expected to be lower in the case of learning algorithms. This also holds for other tasks that employ executive processes, such as language switching. The learning algorithm produces a task-specific mechanism that is utilized in non-target language inhibition. This mechanism may be based on the general inhibitory learning algorithm, but can be expected to be highly specialized, due to the specificity of the task. The learning algorithm framework implies that, during the early stages of language learning, the inhibitory process has not yet become highly specialized, but instead is more similar to processes utilized in other tasks, whereas in fluent bilinguals, correlations between language switching and inhibitory tasks can be expected to be lower.

Let us next focus more closely on the possible functional layout of the second-order functions (SOFs) produced by the learning algorithms. We may consider a SOF as a connectionist network that is optimized for a task though forming new nodes and modulating the connection strengths between the nodes. We may presume that while some nodes of a SOF are unique to that SOF, other nodes engage more general processes (such as motor response inhibition or focusing attention), some of which can be realized as simple algorithms. Thus, even if two tasks (say, Simon and Flanker) both partly relied on motor response inhibition, this sub-process would play a different role in the Simon-SOF and in the Flanker-SOF. This is illustrated in Figure 5, which represents the connectionist layouts of two possible SOFs, one learnt for Flanker and the other for Simon.

My goal here has been to speculate how task-specific cognitive processes might be realized, and why they might even be the optimal strategy for the brain to adapt to tasks. The proposal is not to be considered as a full-blown theory, but instead as a framework that could stimulate further research. It is worthwhile noting that the proposed account of task-specific executive mechanisms is compatible with the brain imaging data indicating that same anatomical regions are utilized in both executive and language switching tasks. Each learning algorithm would be realized in a collection of neurons, which have a specific location in the brain. We may presume that also the secondary, task-specific functions would be realized in roughly the same brain regions, though they would be functionally partly separate.

### 5.3. Directions for future research

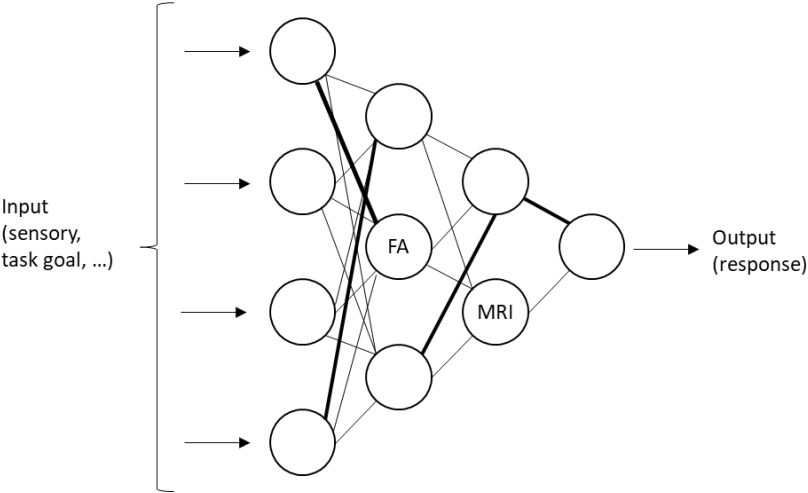
I have argued that behavioral studies address the associations between language switching and EF more directly than brain imaging studies. There have, however, been surprisingly few such behavioral studies, and they have yielded inconsistent results. In many of the previous studies, switch cost asymmetry is considered as evidence of inhibition. Asymmetry, however, can be considered as indirect evidence of inhibition at best, and is based on the IC model's predictions, which can be questioned. Future studies should directly address the associations between language switching and EF. It would be useful for these studies to include a range of executive tasks instead of just one, and to utilize EF tasks that have high convergent validity.

This would help avoiding that that possible inconsistent correlations between language switching and different EF tasks is not due to lack of convergence among the EF tasks themselves. It would also be useful to control for L2 proficiency, as it can be expected that associations between language switching and EF are stronger for those who have recently started learning a new language compared to more advanced learners (in whom the language switching mechanism have become highly specialized processes). Another aspect worthwhile considering would be structural similarity between the language switching and EF tasks: according to the task-specificity account, associations can be expected to be higher between structurally similar language switching and general EF tasks, and lower for structurally dissimilar tasks.

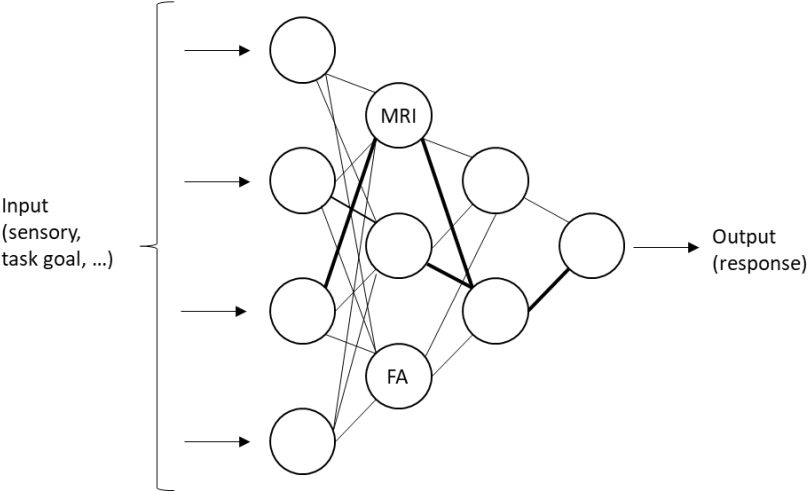
One particularly interesting approach to test for task- or domain-specificity of executive functions would be in clinical samples. Earlier research in aphasic patients indicates dissociations between linguistic and non-linguistic tasks, but they have not directly examined language switching. If language switching relied on general EF, a lesion resulting in weakened language switching performance should also be reflected as weaker general EF performance. Again, if failures in language control occurred in the presence of (near-) intact general EF, this would suggest domain-specificity. Relatedly, transcranial magnetic stimulation (TMS) could be utilized in healthy subjects to cause temporary lesions in the executive system to investigate possible dissociations between the two domains.

Figure 5. Possible SOFs for Flanker (a) and Simon (b). The learning algorithm (in this case, the inhibition algorithm; not represented here) learns a different optimization for both tasks by way of creating new nodes and modulating connection strengths. Basic processes such as Focusing Attention (FA) and Motor Response Inhibition (MRI) may be utilized as a part of the network, but they contribute to a different overall functional layout of the SOF. Line thickness indicates connection strength.

a) SOF-Flanker



b) SOF-Simon





## 6. Conclusions

The Bilingual Executive Advantage hypothesis holds that bilinguals outperform monolinguals on executive functions, due to vast practice in behaviors such as language switching and mixing. Despite promising early findings, recent meta-analyses do not support the view. In this thesis, I have distinguished between two prerequisites for bilingual executive training to occur: the hypothesis that language switching and mixing engages domain-general executive functions, and the hypothesis that these mechanisms can be trained through practice. The results of the present thesis did not support the Training Hypothesis and provided only weak support for the Engagement Hypothesis. I have argued that executive control mechanisms may be highly task-specific, which would explain both why the convergent validity of executive tasks is low, and why bilingual language switching does not consistently correlate with executive performance.

## References

- Abutalebi, J., Brambati, S. M., Annoni, J.-M., Moro, A., Cappa, S. F., & Perani, D. (2007). The neural cost of the auditory perception of language switches: An event-related functional magnetic resonance imaging study in bilinguals. *The Journal of Neuroscience*, *27*(50), 13762–13769. <https://doi.org/10.1523/JNEUROSCI.3294-07.2007>
- Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, *23*(4), 557–582. <https://doi.org/10.1080/01690960801920602>
- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 207–245. <https://doi.org/10.3102/0034654310368803>
- Bexkens, A., Van Den Wildenberg, W. P. M., & Tijms, J. (2015). Rapid Automatized Naming in children with dyslexia: Is inhibitory control involved? *Dyslexia*, *21*(3), 212–234. <https://doi.org/10.1002/dys.1487>
- 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>
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. <https://doi.org/10.1111/j.1467-8624.2007.01019.x>
- Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, *25*(5), 568–585. <https://doi.org/10.1080/20445911.2013.792822>
- Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Theory and Methodology of Frontal and Executive Function* (pp. 81–116). Hove, UK: Psychology Press.
- Calabria, M., Branzi, F. M., Marne, P., Hernández, M., & Costa, A. (2013). 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, F. M., & 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>
- Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., ... Price, C. J. (2006). Language control in the bilingual brain. *Science*, *312*(5779), 1537–40. <https://doi.org/10.1126/science.1127761>
- Dash, T., & Kar, B. R. (2014). Bilingual language control and general purpose cognitive control among individuals with bilingual aphasia: Evidence based on negative priming and flanker tasks. *Behavioural Neurology*, *2014*, 679706. <https://doi.org/10.1155/2014/679706>
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*(1), 99–107. <https://doi.org/10.1177/0956797614557866>
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197.

- <https://doi.org/10.1017/S1366728902003012>
- Donnelly, S. (2016). *Re-Examining the bilingual advantage on interference-control and task-switching tasks: A meta-analysis*. CUNY Academic Works. City University of New York. Retrieved from [http://academicworks.cuny.edu/gc\\_etds/762](http://academicworks.cuny.edu/gc_etds/762)
- Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, 60(9), 1227–1245. <https://doi.org/10.1080/17470210600926075>
- Friedman, N. P., Miyake, A., Young, S. E., Defries, J. C., Corley, R. P., & Hewitt, J. K. (2009). Genetic in Origin. *Journal of Experimental Psychology*, 137(2), 201–225. <https://doi.org/10.1037/0096-3445.137.2.201.Individual>
- Gray, T., & Kiran, S. (2016). The relationship between language control and cognitive control in bilingual aphasia. *Bilingualism: Language and Cognition*, 19(3), 1–20. <https://doi.org/10.1017/S1366728915000061>
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67–81. <https://doi.org/10.1017/S1366728998000133>
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 1–16. <https://doi.org/10.1080/20445911.2013.796377>
- Green, D. W., Grogan, A., Crinion, J., Ali, N., Sutton, C., & Price, C. J. (2010). Language control and parallel recovery of language in individuals with aphasia. *Aphasiology*, 24(2), 188–209. <https://doi.org/10.1080/02687030902958316>
- Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *NeuroImage*, 56(4), 2300–2309. <https://doi.org/10.1016/j.neuroimage.2011.03.049>
- Hartanto, A., & 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>
- Hernandez, A. E., Martinez, A., & Kohnert, K. (2000). In search of the language switch: An fMRI study of picture naming in Spanish–English bilinguals. *Brain and Language*, 73(3), 421–431. <https://doi.org/10.1006/brln.1999.2278>
- Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, 18(4), 625–58. <https://doi.org/10.3758/s13423-011-0116-7>
- Jackson, G. M., Swainson, R., Mullin, a, Cunnington, R., & Jackson, S. R. (2004). ERP correlates of a receptive language-switching task. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 57(2), 223–240. <https://doi.org/10.1080/02724980343000198>
- Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), 6829–6833. <https://doi.org/10.1073/pnas.0801268105>
- Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 42(3), 464–480. <https://doi.org/10.3758/s13421-013-0364-z>
- Johnson, H. A., Sawi, O., & Paap, K. R. (2015). Language switching frequency in bilinguals is inconsistently linked to executive functioning. In *Annual*

- Meeting of the Cognitive Neuroscience Society.*
- Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory, 15*(6), 634–646.  
<https://doi.org/10.1080/09658210701457096>
- Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science, 12*(6), 978–990.  
<https://doi.org/10.1111/j.1467-7687.2009.00846.x>
- Karbach, J., & Verhaeghen, P. (2014). Making working memory work. *Psychological Science, 25*(11), 2027–2037.  
<https://doi.org/10.1177/0956797614548725>
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317–324.  
<https://doi.org/10.1016/j.tics.2010.05.002>
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2017). Is bilingualism associated with enhanced executive control? A meta-analysis. In *Abstracts of the 20th Conference of the European Society for Cognitive Psychology* (p. 235). Potsdam.
- Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15*(3), 651–662.  
<https://doi.org/doi:10.1017/S136672891100054X>
- Liu, H. H., Fan, N., Rossi, S., Yao, P. P., & Chen, B. G. (2016). The effect of cognitive flexibility on task switching and language switching. *International Journal of Bilingualism, 20*(5), 563–579.  
<https://doi.org/10.1177/1367006915572400>
- Liu, H. H., Rossi, S., Zhou, H., & Chen, B. (2014). Electrophysiological evidence for domain-general inhibitory control during bilingual language switching. *PLoS ONE, 9*(10).  
<https://doi.org/10.1371/journal.pone.0110887>
- Macizo, P., Bajo, T., & Paolieri, D. (2012). Language switching and language competition. *Second Language Research, 28*(2), 131–149.  
<https://doi.org/10.1177/0267658311434893>
- Magezi, D. A., Khateb, A., Mouthon, M., Spierer, L., & Annoni, J. M. (2012). Cognitive control of language production in bilinguals involves a partly independent process within the domain-general cognitive control network: Evidence from task-switching and electrical brain activity. *Brain and Language, 122*(1), 55–63.  
<https://doi.org/10.1016/j.bandl.2012.04.008>
- McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*(1), 10–17.  
<https://doi.org/10.3758/BF03195736>
- Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.  
<https://doi.org/10.1006/cogp.1999.0734>
- Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience, 7*(1), 1–17.  
<https://doi.org/10.3758/CABN.7.1.1>
- Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn Affect Behav Neurosci, 12*(2), 241–268.  
<https://doi.org/10.3758/s13415-011-0083-5>
- Paap, K. R., Johnson, H. A., & 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, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, 5(962), 1–15.  
<https://doi.org/10.3389/fpsyg.2014.00962>
- Peeters, D., Runnqvist, E., Bertrand, D., & Grainger, J. (2014). Asymmetrical switch costs in bilingual language production induced by reading words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 284–92.  
<https://doi.org/10.1037/a0034060>
- Pellikka, J., Helenius, P., Mäkelä, J. P., & Lehtonen, M. (2015). Context affects L1 but not L2 during bilingual word recognition: An MEG study. *Brain and Language*, 142, 8–17.  
<https://doi.org/10.1016/j.bandl.2015.01.006>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, 88(5), 879–903.  
<https://doi.org/10.1037/0021-9010.88.5.879>
- Price, C. J., Green, D. W., & Von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain*, 122(12), 2221–2235.  
<https://doi.org/10.1093/brain/122.12.2221>
- Prior, A., & Gollan, T. H. (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>
- Raab, D. H. (1962). Division of psychology: Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*. Blackwell Publishing Ltd.  
<https://doi.org/10.1111/j.2164-0947.1962.tb01433.x>
- Rodriguez-Fornells, A., De Diego Balaguer, R., & Münte, T. F. (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, U. M., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, 3(388), 1–15.  
<https://doi.org/10.3389/fpsyg.2011.00388>
- 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(6), 1477–1491. <https://doi.org/10.1037/0278-7393.31.6.1477>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4, 1–32.  
[https://doi.org/10.1007/978-1-4419-1005-9\\_947](https://doi.org/10.1007/978-1-4419-1005-9_947)
- Soveri, A., Rodriguez-Fornells, A., & 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>
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Oxford: Oxford UP.
- van Heuven, W. J. B., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews*, 64(1), 104–122.  
<https://doi.org/10.1016/j.brainresrev.2010.03.002>
- Van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language

- conflict in the bilingual brain. *Cerebral Cortex*, 18(11), 2706–2716.  
<https://doi.org/10.1093/cercor/bhn030>
- Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & 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>
- von Studnitz, R. E., & Green, D. W. (2002). The cost of switching language in a semantic categorization task. *Bilingualism: Language and Cognition*, 5(3), 241–251.  
<https://doi.org/10.1017/S1366728902003036>
- Wang, Y., Kuhl, P. K., Chen, C., & Dong, Q. (2009). Sustained and transient language control in the bilingual brain. *NeuroImage*, 47(1), 414–422.  
<https://doi.org/10.1016/j.neuroimage.2008.12.055>
- Weissberger, G. H., Wierenga, C. E., Bondi, M. W., & Gollan, T. H. (2012). Partially overlapping mechanisms of language and task control in young and older bilinguals. *Psychology and Aging*, 27(4), 959–74.  
<https://doi.org/10.1037/a0028281>

