



No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost

Kenneth R. Paap, Hunter A. Myuz, Regina T. Anders, Morgan F. Bockelman, Roman Mikulinsky & Oliver M. Sawi

To cite this article: Kenneth R. Paap, Hunter A. Myuz, Regina T. Anders, Morgan F. Bockelman, Roman Mikulinsky & Oliver M. Sawi (2016): No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost, *Journal of Cognitive Psychology*, DOI: [10.1080/20445911.2016.1248436](https://doi.org/10.1080/20445911.2016.1248436)

To link to this article: <http://dx.doi.org/10.1080/20445911.2016.1248436>



Published online: 20 Oct 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost

Kenneth R. Paap^a, Hunter A. Myuz^b, Regina T. Anders^a, Morgan F. Bockelman^a, Roman Mikulinsky^a and Oliver M. Sawi^c

^aDepartment of Psychology, San Francisco State University, San Francisco, CA, USA; ^bDepartment of Psychology, New Mexico State University, Las Cruces, NM, USA; ^cDepartment of Psychology, University of Connecticut, Storrs, CT, USA

ABSTRACT

Participants completed three cued-switching tasks, responded to two category-fluency probes, two letter-fluency probes, and two probes to alternate between two targets. Correlations across the three cued-switching tasks were significant for both switching costs and mixing costs. The bilingual advantage hypothesis was tested both by forming language groups and treating bilingualism as a continuous variable. No bilingual advantages were observed. In verbal-fluency monolinguals generated more correct responses but the bilingual disadvantage on the category task was not reduced in the letter-fluency scores. The bilingual disadvantage was eliminated when the groups were matched on vocabulary size. The verbal-fluency measures obtained when participants alternated between targets weakly correlated with the switching-costs obtained in the cued-switching tasks.

ARTICLE HISTORY

Received 16 March 2016
Accepted 7 October 2016

KEYWORDS

Language switching;
executive function;
bilingualism; switching cost;
verbal fluency

Introduction

The Bilingual advantage in EF hypothesis

More than 100 published articles have tested the hypothesis that bilingualism causes an enhancement in general executive functioning (EF). The hypothesis is predicated on the assumption that coordinating two languages requires extensive use of general-purpose executive functions. Although many of these articles report statistically better performance for bilinguals compared to monolinguals, recent reviews have drawn attention to alternative explanations for these positive results and have suggested that the bilingual advantages may not be real (Hilchey, Saint-Aubin, & Klein, 2015; Paap, Johnson, & Sawi, 2014, 2015, 2016). The bilingual advantage hypothesis is difficult to test for many reasons and one is the complexity of the EF construct, potentially explaining some of the inconsistencies in the findings.

EFs consist of a set of general-purpose control processes that are central to the self-regulation of thoughts and behaviours and that are instrumental to accomplishing goals. Research on EF has often focussed on the three components initially

identified by Miyake et al. (2000) using latent variable analyses: updating, switching,¹ and inhibitory control. Inhibitory control was inferred from performance measures in three different tasks that all involve competition and therefore require some type of conflict resolution such as the inhibition of a prepotent response. Likewise, a general switching ability was inferred from performance on three different tasks that frequently required participants to switch from one task (e.g. judgments about colour) to another (e.g. judgments about shape). The third latent variable—updating of working memory representations—requires monitoring and coding incoming information for task relevance and then appropriately revising the information held in working memory. In Miyake et al. (2000) each of the three observed measures significantly loaded on the expected latent variable establishing that these three EFs can be considered as separate abilities. Furthermore, at the higher level of the analysis, the three latent variables are also correlated with one another and this is consistent with the assumption that the latent variables are components of a common EF ability.

CONTACT Kenneth R. Paap  kenp@sfsu.edu

¹Miyake et al. use the term “shifting”, but we will use “switching” in deference to the vast literature on task switching.

Miyake and Friedman (2012) now favour a variation on the simple hierarchical model described above. They compared the fit of the simple hierarchical model to a more complex second-order (“nested”) model where the nine observed measures are allowed to load on common EF and the three latent variables compete in accounting for the remaining variance. The best solution for the second-order model resulted in all nine measures loading on the common EF and with only two of the nested components (updating and switching) still making unique contributions. Putting this together, the model supports a theory of a general EF ability with separate updating and switching components and an inhibition component that is not separable, but moderately linked to general EF ability. This analysis led Miyake and Friedman (2012) to conclude that EF has both *unity* (a common EF) and *diversity* (additional specific abilities associated with switching and updating).

There are other frameworks, also based on latent variable analyses, that are somewhat different and it may be prudent to not commit to a specific model as our understanding of EF is evolving. Using similar latent variable methods Engelhardt, Briley, Mann, Harden, and Tucker-Drob (2015) used data from 505 third- through eighth-graders and 12 different tasks to test five different models, but the second-order *unity–diversity* model favoured by Miyake and Friedman was not among those evaluated. The best-fitting model includes the inhibitory control and switching components, but also separates working memory capacity (WMC) from updating.

Why focus on switching?

For several reasons the focus of this project is on the switching component of EF. First, a comprehensive review of the bilingual advantage hypothesis for switching has been neglected whereas there have been recent reviews or meta-analyses of the other components. Donnelly, Brooks, and Homer (2015) reported a meta-analysis of 73 comparisons with estimates of global RTs (usually taken as a measuring of monitoring, $d = .43$) and interference costs (usually taken as a measure of inhibitory control, $d = .29$). The meta-analysis reported by de Bruin, Treccani, and Della Sala (2015) also focused on interference scores (although not exclusively) and reported an overall effect size of $d = .30$ based on 176

comparisons. A third recent meta-analysis (von Bastian, De Simoni, et al., 2015, November) focused only on the updating and working memory component and reported average effect sizes of $d = .02$, $.06$, and $.12$ for children, older adults, and young adults, respectively. The 95% confidence intervals (CIs) included zero for all three age groups. Across these meta-analyses the average effect size is very small. If there is a bias favouring positive results that has been instantiated by researchers placing null or negative results in the proverbial file drawer and if the peer-review process has favoured positive results—then the “missing” tests could readily shift the average effect sizes to zero (see Paap, 2014, for a discussion of the file drawer problem in relation to the bilingual advantage).

Other complications for the EF construct and, therefore, another problem for testing the bilingual advantage in EF hypothesis are the psychometric shortcomings of many of the common measures of inhibitory control and monitoring. Monitoring in the popular nonverbal interference tasks (e.g. flanker, Simon, or spatial-Simon) is typically measured as “global RT”, the mean of a block of trials that randomly mixes congruent and incongruent trials and sometimes as only the mean of the congruent trials. Global RT, like any mean RT, suffers from the task impurity problem (Burgess, 1997). In other words, overall performance in a task reflects not only the process of interest (e.g. inhibitory control), but also all other stages of processing from perceptual encoding through response selection and execution. Thus, an individual displaying a very fast global RT may not have superior EF, but rather may have better perceptual-motor skills.

One approach to filtering out these differences in basic processing is to take the difference in RTs between two tasks that differ only in terms of the process of interest. This, of course, is the logic behind using the differences between congruent and incongruent trials to try to isolate the inhibitory control process. Extending this to the monitoring function, the difference between mean RT on the congruent trials in a block that mixes both congruent and incongruent trials and the mean RT from a pure block of neutral (control) trials is sometimes used as a measure of monitoring (e.g. Paap & Greenberg, 2013). Because individual differences in general processing tend to cancel out in the subtraction, this measure is usually credited with being a more pure measure of the monitoring construct. Although difference RTs are “purer” than single-

mean RTs, they do not completely isolate the process of interest.²

The purity advantage favouring difference scores can run counter to the test–retest reliability of the measures. Paap and Sawi (2016), for example, reported that the reliability of the Simon interference effect (a difference measure) is only $r = .344$ ($p = .002$). Thus, only 12% of the variance in Day 2 Simon effects can be accounted for on the basis of the observed Day 1 effects. The test–retest reliability of the interference scores from the flanker task and other Stroop-like tasks are higher than the reliability of the Simon effect, but are usually substantially lower than the less pure single-mean measures (Paap & Sawi, 2016).

Is the test–retest reliability for switching ability adequate? Both of the standard measures are difference scores. One, switching cost, is the difference in mean RT between switch trials and repeat trials in a block where the task is randomly precued on each trial. The other, mixing cost, is the difference in mean RT between the repeat trials from the mixed block and the mean RT from the single-task blocks. Both of these demonstrated acceptable test–reliability ($r = .62$ for switching cost and $r = .65$ for mixing cost) in the Paap and Sawi (2016) study.

A second psychometric issue is that common measures of EF often lack convergent validity, that is, cross-task correlations on measures presumed to reflect the same executive function often have very low and non-significant correlations. Paap and Sawi (2014) reported the cross-task correlations ($N = 120$) for 13 different measures that were derived from four different tasks. The cross-task correlation between the flanker and Simon effects was near zero and non-significant ($r = +.021$, $p = .826$) implying that, at best, only one of these is a valid measure of general inhibitory control. The low levels of convergent validity reported by Paap and Sawi (2014) are consistent with most of the literature they reviewed although there is one case of a modest ($r = +.21$) cross-task correlation between Simon and flanker effects (Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010). To make matters worse, two versions of the same task sometimes fail to correlate. Salthouse (2010) had 265 participants complete both the letter version (Eriksen & Eriksen, 1974) and the arrow version (Stoffels & van der

Molen, 1988) of the classic flanker task and observed a correlation of only $r = +.03$ between the two flanker effects.

The research establishing convergent validity for measures of switching is not as plentiful, but comparatively encouraging. For example, Miyake and Friedman (2012) showed that the switch-cost measures from three different cued-switching tasks strongly loaded on the shifting factor, and the shifting factor was separable from the updating function. Similarly, Klauer et al. (2010) used a set of cued-switching tasks and reported significant cross-task correlations for two of the three tasks, both of which strongly loaded on the switching factor, and the switching factor was related to (but separable from) both the inhibitory control and WMC factors.

In summary, switching ability is the focus of this project because the bilingual advantage hypothesis has not been systematically reviewed for this component of EF, the typical measures of switch costs and mixing costs enjoy higher test–retest reliability than the difference scores used to test for bilingual advantages in inhibitory control and monitoring, and it appears likely that measures of switch costs enjoy higher and more consistent levels of convergent validity. The latter is highly desirable because we still adhere to the recommendation of Paap and Greenberg (2013) that a compelling demonstration of bilingual advantages in EF require statistically significant advantages in different measures derived from at least two different tasks and that those measures must also show convergent validity.

Language switching and general task switching

An aspect of language control unique to multilingualism is the need to switch from one language to another. For example, bilinguals switch languages to accommodate the languages spoken by their conversational partners. Most theories assume that language switching involves the selection of a new *language-task schema*, for example, a *Speak English* schema. An important question is the degree to which switching languages involves the same EFs (and the same neural circuits) that are used for general task switching.

²Paap and Sawi (2016) present a detailed description of this failure to isolate within the context of Miller and Ulrich's (2013) individual differences in RT model.

The overlap hypothesis

Does switching back and forth between two nonverbal tasks involve the same control mechanism as switching back and forth between English and Spanish? If it does, then bilinguals, especially those who switch languages frequently, should have an advantage in general task switching compared to otherwise similar monolinguals who do not receive this extra practice. In its strongest form one might assume that language switching requires the same control mechanism used for general task switching regardless of the direction of the switch (L1 to L2 or the reverse) and the relative proficiencies in the two languages. This assumption will be referred to as the overlap hypothesis.

The independence hypothesis

However, there is substantial behavioural evidence that language switching is not subsidiary to general task switching and that the functional overlap between the two types of switching is minimal. This alternative will be referred to as the independence hypothesis. The independence hypothesis is consistent with research showing that language-switching performance does not correlate with performance on nonverbal switching tasks (Calabria, Branzi, Marne, Hernández, & Costa, 2015; Calabria, Hernández, Branzi, & Costa, 2011; Yim & Bialystok, 2012), that nonverbal task switching shows a much greater decline with age (Calabria et al., 2015) and that general switching abilities can be severely impaired while language switching is not (Weissberger, Wierenga, Bondi, & Gollan, 2012).

On the other hand, Festman et al. (Festman & Münte, 2012; Festman, Rodríguez-Fornells, & Münte, 2010) reported an interesting association between language-switching errors and performance on common measures of inhibitory control (e.g. the flanker effect) or switching (e.g. Wisconsin Card Sorting Task). In these studies participants are cued to switch languages in alternating runs (viz., RRGRRGG ... where R is picture naming in Russian and G is naming in German) and are partitioned into good and poor switchers on the basis of the number of “errors of interference”, that is, responses that are semantically correct, but in the wrong language.

Furthermore, De Baene, Duyck, Brass, and Carreiras (2015) using fMRI showed that a highly similar fronto-parietal network was involved in language switching and nonverbal task switching within a group of 32 highly proficient and early

Basque-Spanish speakers who were also proficient in English. It would be useful to know how precisely and completely the neural circuits match. Because the disjunctive analyses identified regions involved in language switching, but not general task switching and the reverse, the overlap is not complete. Of course, as discussed by De Baene et al., differences in the two tasks (e.g. response modality and arbitrariness) accounts for some of the disjunction. Turning to the overlapping regions, to what extent are identical neural circuits involved in language switching and general task switching? This is probably a matter of informed conjecture regarding the homogeneity of the millions of neurons within each voxel. Relatedly, the average voxel-by-voxel correlation between the magnitude of the switch-repeat contrasts obtained in the language versus nonverbal tasks was about $r = .50$. Thus, the degree to which voxel activation is modulated by language switching predicts only about 25% of the variance in nonverbal switching. In summary, this excellent study provides compelling evidence for partial overlap in the neural activity underlying language switching and general task switching. Still, the low correlations and dissociations in the behavioural data described earlier motivate one to keep an open mind and not automatically presume that language switching is subsidiary to general task switching. Furthermore, as García-Pentón, Fernández García, Costello, Duñabeitia, and Carreiras (2016) caution, demonstrating overlap in the distributed fronto-parietal network has no direct implications for the bilingual advantage hypothesis. The evidentiary chain would need to show that the switching networks of bilinguals are more efficient than those of monolinguals and, most important, that any neural differences align with behavioural advantages in nonverbal switching costs.

Frequency of language switching

The independence hypothesis would be severely challenged if frequency of language switching (FoLS) was strongly related to bilingual advantages in EF. Indeed, Verreyt, Woumans, Vandelandotte, Szmalec, and Duyck (2016) concluded that the frequency with which bilinguals switch languages is the “key determinant” in potentiating a bilingual advantage in EF. They reported that high-switch bilinguals had smaller Simon and flanker interference effects compared to low-switch bilinguals. This challenge is somewhat muted by two aspects

of the design. First, the sample size was small with 20 high-switch bilinguals and 17 low-switch bilinguals. Second, no monolinguals were tested to validate that the high-switch advantage would carry over to a group of monolinguals.

The latter problem was corrected in a follow-up study³ (Woumans, Ceuleers, Van der Linden, Szmalec, & Duyck, 2015) that included four groups: French monolinguals ($n = 30$), unbalanced Dutch–French bilinguals ($n = 34$), balanced bilinguals ($n = 31$), and interpreters ($n = 28$). As shown parenthetically, the sample sizes were also somewhat larger, but the design is still underpowered by recommended standards. Depending on the comparisons and analyses considered, the newer results may also challenge the independence hypothesis. All three bilingual groups showed smaller Simon effects. This is evidence favouring the bilingual advantage hypothesis in general, but not the more specific variant that FoLS is a key determinant. Given the important role of FoLS in Verreyt et al. (2016) it is surprising that a FoLS measure is not reported in Woumans et al. (2015). Since the unbalanced bilinguals reported using their L2 only 7.7% of the time it is probably safe to assume that they switched languages far less than the balanced bilinguals (35.5% L2 use) and certainly less than the interpreters who in all likelihood are ultra-high switchers. Yet, there were no bilingual-group differences in the Simon effect (incongruent RT—congruent RT) or in the flanker effect. In this regard the second study fails to replicate the first study.

Woumans et al. (2015) do investigate the relationship between language control and cognitive control by correlating a cost-of-switching measure with the various Simon and flanker measures for each group separately. The cost-of-switching measure was derived from a verbal-fluency task that required the bilinguals to alternate languages while producing instances of a semantic category. The cost-of-switching measure is the difference in L1 performance between this alternating condition and a baseline where all the responses are given in L1. Although not explicitly stated, there appear to be 24 tests: two tasks (Simon, flanker) $\times$ two measures of EF (interference score, global RT) $\times$ two types of measures (latency, accuracy) $\times$ three bilingual groups. The cost-of-

switching measure significantly predicted the magnitude of the Simon interference effect, but only for the group of balanced bilinguals. This is not very compelling evidence that a bilingual's ability to switch languages predicts measures of EF derived from nonverbal tasks. Also, although the cost-of-switching measure used by Woumans et al. probably correlates with the FoLS used by Verreyt et al., it is not the same thing. For example, it is likely that scores on the cost-of-switching measure reflect not only the amount of practice in switching languages, but also the relative proficiency in both languages.

In a more direct test of a possible link between the FoLS and measures of EF derived from Simon and flanker tasks; Johnson, Sawi, and Paap (2015) reported analyses similar to Verreyt et al. (2016) on a composite database formed by combining the Simon and flanker data from Paap and Greenberg (2013, Study 3) with that from Paap and Sawi (2014).⁴ For the Simon and flanker data reported by Johnson et al. (2015) there were no significant differences between high-switch bilinguals ($n = 40$), low-switch bilinguals ($n = 56$), and monolinguals ($n = 66$). When the same data were reanalysed using all five rating-scale values (rather than partitioning the bilinguals into high- and low-switchers) Spearman's rho for each measure of EF was non-significant. Likewise, in a study by Yim and Bialystok (2012) FoLS (based on an objective measure rather than self-reports) also had no effect on a measure of inhibitory control.

Note that the low levels of convergent validity typically observed between the Simon and flanker effects comes into play when considering the results of Woumans et al. (2015). With respect to measures of inhibitory control the bilingual advantage occurred with the Simon effect, but not with the flanker effect. Likewise, with respect to global RTs a bilingual advantage occurred in the flanker task, but not the Simon task. These types of dissociations are to be expected if the two tasks are measuring different constructs. Given the assumption that switching-cost measures enjoy greater convergent validity, it makes both theoretical and psychometric sense to test for a relationship between FoLS and switch costs and that is part of the design of the present study. The present study

³Although Verreyt et al. has a publication year of 2016 and Woumans et al. a publication year of 2015, Verreyt et al. was received at the journal a year before Woumans et al.

⁴The rating scale and the derivation of the frequency-of-switching variables are the same as those described later in the results section where we perform a similar analysis of switching costs.

together with the colour–shape switching data obtained by Paap and Greenberg (2013) and Paap and Sawi (2014) will enable the formation of very large database of bilinguals that will enable further exploration of the relationship between frequency-of-switching and switch costs.

Bilingual advantages in switching or mixing costs

Paap, Johnson, and Sawi (2015) credit the Prior and MacWhinney (2010) study with launching a flotilla of

tests for bilingual advantages in cued-switching tasks and discussed seven studies that followed in its wake. Table 1 provides an update of the chronology for the 17 studies and 53 tests for bilingual advantages that used some version of the colour–shape switching task. The bilingual advantage in switching costs reported by Prior and MacWhinney looked like one that should easily replicate given that the estimated effect size was $d = .52$ (with 44 participants in each group) and that the estimated power for a one-tailed test with an alpha equal to .05 was .78. However, as shown in Table 1 the

Table 1. Results of tests for bilingual advantages in colour–shape switching tasks.

Article	<i>n</i> per group	Age	Switch cost	Mixing cost	Global cost
Prior and MacWhinney (2010)					
Various bilinguals	44	19.5	B+	ns	
Prior and Gollan (2011)					
Spanish–English bilinguals	41	20.0	B+	ns	
Mandarin–English bilinguals	43	19.4	ns	ns	
Tare and Linck (2011)					
Various bilinguals	35	27.7	ns	ns	
Barac and Bialystok (2012)					
Chinese–English bilinguals	30	6.0			B+
French–English bilinguals	28	6.2			B+
Spanish–English bilinguals	20	6.2			B+
Gold, Kim, Johnson, Kryscio, and Smith (2013)					
Exp. 1: various bilinguals	15	63.3			B+
Exp. 2: various older bilinguals	20	older			ns
Exp. 2: various young bilinguals	20	younger			ns
Prior and Gollan (2013)					
Spanish–English bilinguals	30	20.3	ns	ns	
Hebrew–English bilinguals	6	25.0	ns	ns	
Mandarin–English bilinguals	16	20.0	ns	ns	
Paap and Greenberg (2013)					
Study 1: various bilinguals	34	US	ns	ns	
Study 2: various bilinguals	36	US	ns	ns	
Study 3: various bilinguals	52	US	ns	ns	
Hernández, Martín, Barceló, and Costa (2013)					
Experiment 1: Spanish–Catalan bilinguals	87	20.6	ns		
Experiment 2: Spanish–Catalan bilinguals	20	20.4	ns		
Experiment 3: Spanish–Catalan bilinguals	38	19.9	ns		
Rodríguez-Pujada et al. (2013)					
Catalan–Spanish bilinguals	18	23.1	ns		
Paap and Sawi (2014)					
Day 1: various bilinguals	58	US	ns	ns	
Day 2: various bilinguals	39	US	ns	ns	
Mor, Yitzhaki-Amsalem, and Prior (2014)					
Hebrew–English bilinguals	20	24.8	ns	ns	
Moradzadeh, Blumenthal, and Wiseheart (2014)					
Non-musicians: various bilinguals	36	21.5	ns	ns	
Musicians: various bilinguals	36	22.5	ns	ns	
Wiseheart, Viswanathan, and Bialystok (2014)					
Various bilinguals	31	19.2	ns	B+	
Houtzager, Lowie, and de Bot (2014)					
Middle-aged: Dutch–Frisian bilinguals	25	46.0	ns	ns	
Elderly: Dutch–Frisian bilinguals	25	73.2	B+	ns	
de Bruin, Bak, and Della Sala (2015)					
Older active Gaelic–English bilinguals	28	71.9	ns	ns	
Older inactive Gaelic–English bilinguals	24	70.5	ns	ns	
Shulley and Shake (2016)					
Various bilinguals	58	21.3	ns		
Branzi, Calabria, Gade, Fuentes, and Costa (2016)					
Exp. 2: Spanish–Catalan bilinguals	91	US	ns		

Note: B+ = significant bilingual advantage at $p < .05$; US = university students; ns = non-significant.

bilingual advantage with young adult populations obtained by Prior and MacWhinney has been replicated only once (Prior & Gollan, 2011).⁵ Despite the recent failures to replicate, Tao, Taft, and Gollan (2015) recently commented that “the bilingual advantage in nonlinguistic task switching appears to be more replicable than advantages in other aspects of cognitive functioning ...” (p. 532). This is simply not true as explicated below.

The main purpose of Table 1 is to show that the colour–shape switching task yields overwhelmingly null results between groups of bilinguals and monolinguals. However, there are studies using other cued-switching tasks or which do not include a monolingual group that appear to support the hypothesis that bilinguals are advantaged in switching ability. Christoffels, de Haan, Steenbergen, van den Wildenberg, and Colzato (2015) obtained a switch-cost advantage for 17-year-old Dutch–English bilinguals who were enrolled in bilingual education versus bilinguals who were enrolled in a monolingual-education programme. The baseline group is not monolingual and in terms of English (L2) proficiency their L2 proficiency is 85% of the bilingual-education group on both subjective rating of proficiency in speaking and an objective measure of vocabulary. If one wants to open the doors to comparisons that do not include “pure” monolinguals, then one must also permit entry to von Bastian, Sousa, and Gade (2015) who used a *k*-means clustering procedure to consider three continuous dimensions (age of L2 acquisition, non-L1 usage, and the L1/L2 proficiency ratio) and to create groups that minimised within-group variance and maximised between-group variance. Three dimensions of bilingual experience that have been hypothesised to play important roles in fostering bilingual advantages in EF were used. There were no bilingual advantages in switching costs for colour–shape, animacy–size, or parity–magnitude tasks. In contrast to the von Bastian et al.’s results, Marzecová et al. (2013) found a bilingual advantage in switch costs (for participants 27 years old) when switching to a gender decision about pictures of faces, but no differences when switching to an age decision. This finding for one dimension (gender) of one task (gender/age) pales in comparison to the consistently null results

across three tasks reported by von Bastian, De Simoni, et al. (2015).

Switching categories in verbal-fluency tasks

Our primary focus is on the switching costs and mixing costs derived from the three cued-switching tasks that have established convergent validity. In addition to the cued-switching tasks, participants in the present study were also required to complete: (1) standard verbal-fluency tasks with either letter targets or category targets and (2) variants of these tasks where responses alternated between two letter targets or category targets. Including these verbal-fluency tasks has two purposes. First is to test the assumption that performance on these alternating-targets tasks provides a measure of general switching ability comparable to switching costs derived from cued-switching tasks. A second purpose is to further investigate the consequences of these four common assumptions: (1) letter fluency requires more EF than category fluency, (2) bilinguals have enhanced EF, (3) category fluency makes greater demands on lexical-access skills than letter fluency, and (4) bilinguals have poorer lexical-access skills. In combination these four assumptions predict a specific pattern of Group $\times$ Task interaction, namely, that the bilingual disadvantage in category fluency should be reduced, eliminated, or possibly even reversed in a letter-fluency task (Bialystok, Craik, & Luk, 2008a; Gollan, Montoya, & Werner, 2002; Luo, Luk, & Bialystok, 2010; Tao et al., 2015). An example is shown in Figure 1 which was constructed from the means presented in Rosselli et al. (2000) for older participants when tested in Spanish. The Group $\times$ Task interaction was statistically significant and shows a bilingual disadvantage in the category task presumably due to their inferior lexical-access skills. However, when their presumably superior EF can be brought to bear on performance in the letter task they “caught up” and both groups produced the same number of correct responses.

As shown in Table 2 a small number of tests have resulted in a significant Group $\times$ Task interaction that shows larger group differences in the category task than in the letter task. There are also studies

⁵In a recent review (Paap et al., 2015) we questioned the bilingual advantage obtained with Prior and Gollan’s Spanish–English bilinguals because the advantage was restricted to an analysis that transformed absolute RT to a measure of relative speed and used a covariate (parents’ educational level) that was confounded across the language groups. Due to an oversight on our part, we failed to notice and report that the advantage was also obtained for a subset of the Spanish–English bilinguals that were comparable to the monolinguals in parents’ educational level.

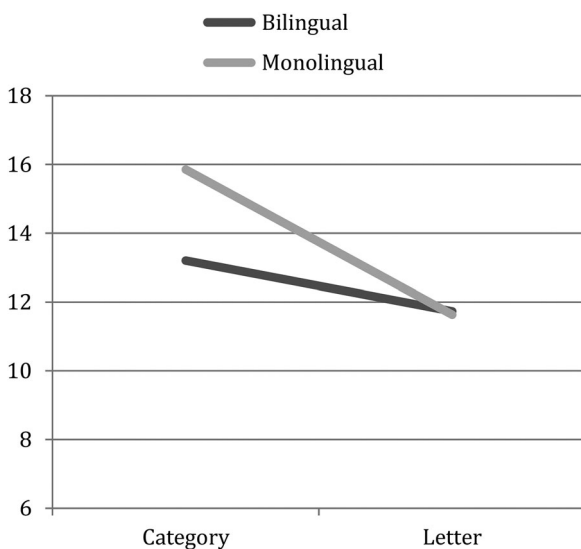


Figure 1. Mean number of correct responses for older bilinguals and monolinguals in both category- and letter-fluency tasks (from Rosselli et al., 2000).

that have not conducted or reported the outcome of the Group $\times$ Task interaction (indicated with a “?” in Table 2). However, most have resulted in a non-significant interaction. Well-versed readers may note that the study of aging Swedish adults published by Ljungberg, Hansson, Andrés, Josefsson, and Nilsson (2013) is not listed in the table. This is because the results of this study are difficult to interpret and categorise. The letter-fluency tasks consisted of a single trial where the target is the letter A. More problematical is that the only “category” target was occupations beginning with the letter B. As the latter actually combines the requirements of both letter- and category fluency into one rule it is not a clear case of category fluency. There was a bilingual advantage for the letter A over the mixed condition, but it was also true that the bilinguals were younger, more highly educated, and included

twice as many females. The present study provides another test of this critical Group $\times$ Task interaction.

Method

Participants

The 236 participants were San Francisco State University undergraduate students who earned extra credit in a psychology course for their participation. Table 3 compares the monolinguals to the bilinguals on self-ratings of English proficiency, correct responses on the multilingual naming task (MINT) (Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2011), age, and number of correct responses on the Ravens test of general fluid intelligence. As shown in the table, the two language groups do not differ with respect to age or Ravens score.

As expected, there is a statistically significant (but small) monolingual advantage in both measures of English proficiency. Perhaps more important, the mean rating for the bilinguals show that, on average, they rate their English ($M = 6.2$) as highly proficient. A rating of “6” on our 7-point scale (see Paap & Greenberg, 2013, for details) is labelled *as fluent as a native speaker* and the top value of “7” is labelled *super fluent—more fluent than a typical native speaker*. Consistent with our earlier work we classified a participant as bilingual if they rated their proficiency in two or more languages as a “4” (*can converse with little difficulty with a native speaker on most everyday topics, but with less fluency than a native speaker*).

To assess the degree of bilingualism there are advantages to classifying a speaker’s languages as L1 and L2, rather than English and other. L1 is operationally defined in this study as the language with the highest rated proficiency (in speaking and

Table 2. Tests of Group $\times$ Task interaction in verbal-fluency tasks.

Dataset	G $\times$ T interaction	Mean age (years)	n per group
Bialystok et al. (2008a)	no	20	24
Bialystok et al. (2008a)	no	68	24
Bialystok, Craik, and Luk (2008b, Study 1)	no	20	24
Bialystok et al. (2008b, Study 2)	no	21	16, 50
Gollan et al. (2002)	yes	20	29, 27
Kousaie, Sheppard, Lemieux, Monetta, and Taler (2014)	no	21	30, 51
Kousaie et al. (2014)	no	72	30, 36
Mor et al. (2014)	no	25	20, 25
Vega-Mendoza, West, Sorace, and Bak (2015)	no	22	16, 18
Rosselli et al. (2000, test in English)	yes	62	45, 19
Rosselli et al. (2000, test in Spanish)	yes	62	18, 19
Sandoval et al. (2010)	no	20	24, 30
Tao et al. (2015, Spanish–English)	?	21	80, 60
Tao et al. (2015, Mandarin–English)	?	20	80, 60
Present study	no	22	96, 111

Table 3. Characteristics of bilinguals and monolinguals.

Measure	Bilinguals		Monolinguals		Diff	SE	<i>t</i>	<i>p</i>
	<i>n</i>	Mean	<i>n</i>	Mean				
English proficiency	122	6.27	108	6.65	−0.38	0.10	−3.80	.000
MINT total correct	113	59.74	103	63.05	−3.31	0.57	−5.79	.000
Parent's education	120	3.28	108	3.75	−0.47	0.20	−2.29	.023
Age	120	21.14	108	22.28	−1.14	0.74	−1.56	.121
Raven's matrices	117	8.21	104	8.64	−0.44	0.30	−0.44	.148

Note: Diff = difference between group means, SE = standard error.

listening) independent of age-of-acquisition (AoA). Similarly, L2 is the language the speaker rates second highest in proficiency. If two languages tie, the language with the lowest (AoA) is designated as the L1. Table 4 shows the frequency distribution across the 7-point rating scale for both L1 and L2 by language group. Because L1 proficiencies tend to be high (viz., fluent or super fluent) for both groups, it is the L2 proficiencies at the intermediate scale values that reflect the fuzzy boundary between monolinguals and bilinguals. Note that 25% of the bilinguals rated their L2 as only a “4” (*can converse with little difficulty with a native speaker on most everyday topics, but with less fluency than a native speaker*). Likewise, 20% of the monolinguals rated their L2 as high as a “3” (*can converse with a native speaker only on some topics and with some difficulty*). We constructed the scale to have the face validity of separating speakers who can engage a conversation with little difficulty from those who cannot. Nonetheless, as we have done in our earlier work we also analyse the present results by both forming more distinctive groups and by treating the L2/L1 ratio as a continuous measure of bilingualism.

Of the 122 bilinguals 37 are native speakers of two languages. For the sequential bilinguals 13 acquired English as their native language while 72 were native speakers of a language other than English. The median AoA of L2 for the sequential bilinguals was 6.0 years old. Collectively the bilinguals spoke 21 different languages, but 52% spoke Spanish, 16% Chinese, and 8% Tagalog.

The six correlations among the four measures of English proficiency (self-reported English proficiency

in speaking and listening, self-reported proficiency in reading English, MINT total correct responses, and category fluency) are significant with all p 's $\leq .005$. The largest correlation, $r = +0.53$, is between category fluency and MINT total correct and this makes sense given that they both focus on productive vocabulary. Given that self-rated English proficiency in speaking/listening was used to classify speakers as monolingual and bilingual in the ANOVA analyses it is noteworthy that the correlation between self-rated speaking and listening and MINT performance was $r = +0.49$. The correlation is quite large given the restricted range of English proficiency (viz., 97% of the ratings were 5, 6, or 7).

Correlations less than perfect should not be attributed exclusively to the shortcoming of the self-rating. Although it is certainly true that individual assessments may under- or overestimate true proficiencies their self-ratings may be the only practical method of assessing general speaking and comprehension skills. Fairly quick, objective, and standardised measures such as MINT usually focus on a single aspect of language proficiency (e.g. productive vocabulary), and provide little or no direct information about competencies in syntax, pragmatics, sentence comprehension, or the ability to produce a fluent sentence. Although this study and earlier ones that use bilinguals who speak a wide variety of languages would clearly benefit from additional objective measures of the non-English languages, it is fair to recall Marian, Blumenfeldt, and Kaushanskaya's (2007) conclusion that self-reported proficiency “is an effective, efficient, valid, and reliable tool for assessing bilingual

Table 4. Frequency distribution of L1 and L2 proficiencies for bilinguals and monolinguals.

		Rating of proficiency in speaking and listening							
	<i>n</i>	0	1	2	3	4	5	6	7
Bilinguals	122								
L1		0	0	0	0	1	2	40	79
L2		0	0	0	0	31	40	39	12
Monolinguals	108								
L1		0	0	0	0	0	3	32	73
L2		52	12	22	22	0	0	0	0

Table 5. Correlations between parents' educational level and switching and mixing costs.

	Colour-Shape		Letter-Digit		Animacy-Size	
	Switching	Mixing	Switching	Mixing	Switching	Mixing
Pearson <i>r</i>	−.041	+.052	+.005	−.076	+.008	−.022
Sig. (2-tailed)	.547	.445	.939	.265	.905	.743
<i>N</i>	216	216	216	216	221	221

status" ($p = .960$). The conclusion was based on the correlations between self-reported measures of both L1 and L2 proficiency and eight standardised behavioural measures of proficiency. In our previous work (Paap & Greenberg, 2013; Paap & Liu, 2014; Paap & Sawi, 2014) we have shown that our self-rated English proficiency scale significantly predicts performance in: a sentence comprehension task requiring resolution of lexical ambiguity, judging if sentences contain a semantic anomaly, judging if sentences contain a syntactic error, judging if letters are English words or nonwords, and category fluency.

As shown in Table 3 the parent's educational level (PED) is significantly lower for the bilinguals. This confound would be important if PED was associated with better scores (e.g. smaller switching or mixing costs) on measures of EF. In our earlier work (Paap & Greenberg, 2013; Paap & Sawi, 2014) using the same participant pool and recruiting methods, we found a similar bilingual disadvantage on PED, but the correlations between PED and measures of monitoring, inhibitory control, and switching were always non-significant and often near zero. This is also the case in the present study and as shown in Table 5 the correlations between PED and the mixing and switching costs for the cued-switching tasks are all non-significant, less than $r = .08$, and indeed sometimes in the unanticipated direction (i.e. positive).⁶

Although we did not explicitly ask about immigrant status, on the basis of ongoing work that recruits from the same student population we know that about one-third of our bilinguals immigrated with their families to the United States as children, but only a few did so on the own initiative as adults. Immigrant status is often associated with higher intelligence and more generally with a "healthy migrant effect" (Bak, 2015; Fuller-Thomson & Kuh, 2014; Kirk, Fiala, Scott-Brown, & Kempe, 2014). If the healthy immigrant effect generalises to EF, then we would expect the group with a disproportionate number of immigrants to have better switching ability. Thus, immigrant status cannot explain why bilinguals showed no switching

advantages, but the unknown effects of immigration and culture are a limitation of the present study.

Cued-switching tasks

Colour-Shape switching task

We employed a version of the colour-shape switching task that is very similar to that used by Prior and MacWhinney (2010) and identical to the task used in Paap and Greenberg (2013) and Paap and Sawi (2014). During a pure block of colour decisions participants responded with two fingers of the left hand to indicate if the target was blue or red. Throughout a pure block of shape decisions participants responded with two fingers of the right hand to indicate if the target was a circle or a triangle. In the mixed blocks, a precue (either a rainbow or a black circle embedded in a black triangle) was presented 250 ms before the target on each trial to signal whether colour or shape was the relevant dimension. Thus, each trial in a mixed block was either a repeat (same dimension as the previous trial) or a switch (different from the previous trial). The task consisted of six blocks. The first block of 16 trials was pure colour with each of the targets appearing in random order. The second block was pure shape. Following Block 2 the mixed task was introduced with detailed instructions regarding the use of the precue to signal whether a colour or shape judgment would be required on each specific trial. Each of the four mixed blocks started with two buffer trials that were not analysed. Block 3 was considered practice and consisted of 18 trials including the two buffers. Blocks 4, 5, and 6 each consisted of 50 trials (including the two buffers). A single random order was used for every participant. Each of the four targets appeared 36 times across Blocks 4 to 6 and there were 72 repeat trials and 72 switch trials. Single-task blocks of 36 trials each preceded the three mixed blocks of 48 trials each. In each mixed block half the trials were repeat trials and half were switch trials.

⁶It is possible that some other measure of childhood SES is also lower in our bilingual group and that this hypothetical other measure might correlate with mixing or switching costs.

Digit–letter and semantic-category tasks

The other two cued-switching tasks were modelled after those used by Friedman et al. (2008) and Miyake and Friedman (2012) who also used a colour-shape switching task. The letter–number task presented a digit/letter pair on each trial and participants are cued to either make a consonant–vowel or an even–odd decision. The animacy–size task presented a printed noun on each trial and participants are cued to either make a size decision (smaller or larger than a soccer ball) or a living–non-living decision. Consistent with Friedman et al. (2008) the 16 nouns were: *table, bicycle, coat, cloud, pebble, knob, marble, snowflake, shark, lion, oak alligator, mushroom, sparrow, goldfish, and lizard*. The structures of the single-task and mixed-task blocks were identical to that described for the colour-shape task. These three cued-switching tasks were also used by both Friedman et al. (2008) and Miyake and Friedman (2012) and in both latent variable analyses the loading on the “switching” factor was substantial for each task, ranging from .61 to .73 with a mean of .67.

The verbal-fluency tasks

We also included two switching measures from the D-KEFS battery (Delis, Kaplan, & Kramer, 2001) of EF measures. Both involved verbal fluency as participants were asked to produce as many words as they could in one minute that followed the prescribed rule. In the alternating-letters task participants had to alternate between words beginning with two target letters. A first trial required alternation between *S* and *T* and a second trial between *C* and *N*. In the alternating-categories task participants had to alternate between two semantic categories. One trial alternated between *animals* and *countries* while a second between *fruits* and *furniture*. Two trials of simple letter fluency (words beginning with *F* and words beginning with *A*) and two trials of simple category fluency (musical instruments and vegetables) were also included. These single-target conditions are sometimes used as a baseline for inferring the cost of having to switch (alternate).

Results

Single-task baseline performance

RTs and proportion correct were averaged across the pure colour and pure shape blocks to compute a

baseline for single-task performance in each of the three cued-switching tasks. The only significant group difference was in proportion correct on the animacy–size task, $t(210) = .93$, $p = .004$. The mean for monolinguals ($M = .927$, $SD = .05$) was higher than that for bilinguals ($M = .905$, $SD = .06$). This monolingual advantage in the accuracy of animacy–size judgments may indicate that the animacy–size task is sensitive to small differences in test-language proficiency. Accuracy in these single-task blocks significantly correlated with both rated proficiency ($r = +0.18$, $p = .009$) and MINT scores ($r = +.47$, $p < .001$), whereas the correlations in the other two tasks are small and non-significant.

Colour-shape task

Table 6 compares bilinguals to monolinguals on measures of switching costs and mixing costs for the three cued-switching tasks. Each row shows the mean for each language group, the group differences, the standard error of the difference, the associated t -statistics, and the exact probability associated with the t value. The last column shows the Bayes Factor (BF) (Rouder, Speckman, Sun, Morey, & Iverson, 2009) which is the ratio of the probability of the null given the data to the probability of the alternative hypothesis given the data. The output of Rouder’s calculator places the more likely hypothesis in the numerator with an indication of whether it is the null (an “n” in the last column of Table 6) or the alternative (an “a” in the last column) that is more likely given the evidence. BFs between 0 and 3 are often characterised as having only an “anecdotal” level of support for the more likely hypothesis while evidence increases from moderate to strong as the BFs further increase from 3 to 10 (Jeffreys, 1961).

Switching costs

The most common measure of switch cost is the difference in mean RTs between the repeat and switch trials. As shown in Table 6 the groups do not significantly differ, nor is there a difference in switch costs computed from proportion correct. Because the absence of a significant difference in proportion correct does not assure the absence of a subtle speed–accuracy trade-off the table also shows switch costs based on efficiency scores (ES) which were calculated by dividing the condition mean RTs by the proportion correct (Christie & Klein, 1995; Townsend & Ashby, 1983). The

Table 6. Group differences in switching and mixing costs.

Task	Biling.	Mono.	Diff	SE	<i>t</i>	<i>p</i>	BF	H
Colour–Shape Task	<i>n</i> = 110	<i>n</i> = 95						
Switch cost RT	223	216	7	20	0.34	.74	6.2	n
Switch cost PC	.022	.030	–.007	.007	–1.03	.31	4.0	n
Switch cost ES	283	275	8	32	0.24	.81	6.4	n
Mix cost RT	348	261	87	42	2.06	.04	1.1	a
Mix cost PC	–.026	–.017	.009	–.014	–0.64	.52	5.5	n
Mix cost ES	401	297	104	58	1.79	.07	1.5	n
Letter–Digit Task	<i>n</i> = 114	<i>n</i> = 102						
Switch cost RT	306	320	–14	23	–0.60	.55	5.7	n
Switch cost PC	.034	.037	–.003	.007	–0.53	.60	5.9	n
Switch cost ES	365	386	–21	28	–0.76	.45	5.1	n
Mix cost RT	70	133	–63	36	–1.76	.08	1.6	n
Mix cost PC	.012	.009	.003	.008	0.40	.69	6.2	n
Mix cost ES	401	297	104	58	1.79	.07	1.5	n
Animacy–Size Task	<i>n</i> = 110	<i>n</i> = 102						
Switch cost RT	141	122	19	19	0.99	.32	4.2	n
Switch cost PC	.012	.004	.008	.005	1.49	.14	2.4	n
Switch cost ES	180	133	47	34	1.40	.16	2.6	n
Mix cost RT	2	–21	23	38	0.61	.54	5.6	n
Mix cost PC	.050	.030	.020	.007	2.72	.01	4.6	a
Mix cost ES	–52	–57	5	49	0.10	.92	6.6	n

Note: Diff = group difference, BF = Bayes Factor, H = hypothesis favoured by BF, n = null hypothesis, a = alternative hypothesis.

absence of significant differences across all of the switching-cost measures and the moderately high BF values provides no support for the hypothesis that young adult bilinguals have enhanced switching ability.

Mixing costs

In contrast to the results of switching costs, there was a significant monolingual advantage in mixing costs, although accompanied by a very modest BF value of 1.1. Furthermore, the difference was no longer significant in the ES mixing-cost measure.

Gapped groups

All of the *t*-tests shown in Table 6 were repeated after deleting monolinguals who rated their L2 proficiency as a “3” and after deleting bilinguals who rated their L2 proficiency as a “4”. These analyses were based on 82 bilinguals and 77 monolinguals. There were no significant differences between the language groups, all *p*'s > .15. Thus, the only difference in statistical outcome was that the monolingual advantage in mixing-cost RT was no longer significant in the gapped-group analysis, *t*(157) = 1.38, *p* = .171.

Letter–digit task

Switching costs

As shown in Table 6 there were no significant differences between the language groups in switching-cost accuracy, RT, or ES. The BFs for the three

switching-cost measures are all between 5 and 6, reflecting moderately strong support for the null hypothesis.

Mixing costs

Likewise there were no significant differences between the language groups in mixing-cost accuracy, RT, or ES. The mixing costs for RT and ES do trend toward a bilingual advantage, but the BFs of about 1.5 indicate that the relative evidence favours the null hypothesis at levels that are typically characterised as “anecdotal” or “barely worth mentioning”.

Gapped groups

Eliminating the monolinguals with L2 proficiency of “3” and bilinguals with a “4” leaves samples sizes of 80 and 83, respectively. The outcome of the six statistical comparisons remained the same, no significant differences between the language groups, all *p*'s > .18.

Animacy–size task

Switching cost

As shown in Table 6 there were no significant differences between the language groups in switching-cost accuracy, RT, or ES with BFs of 4.2, 2.4, and 2.6, respectively. Although the BFs consistently favour the null hypothesis, only the switching-cost RTs are of moderate strength.

Mixing cost

The mixing-cost RTs in the animacy-size task are very small and for the monolinguals are actually slightly negative indicating that on average RTs on the repeat trials were faster than those in the pure blocks. This is caused by the combination of two aspects of the task design. First, in comparison to the letter-shape or letter-digit task, the animacy-size judgments are less familiar and more difficult, especially the first time or two through the set of 16 nouns. As participants pointed out to us some *lizards* are bigger than a soccer ball, as are *marble* slabs, but not a *marble* used to play marbles. *Mushrooms* have the defining features of living things, but not many of the characteristic ones. Because the single-task blocks preceded the mixed blocks, they took the brunt of the initial pondering and washed out the costs of mixing. In retrospect it would be far better to provide additional practice or to use the sandwich design employed by Prior and MacWhinney where the single-task blocks were repeated after the mixed blocks. We view our mixing-cost measures in the animacy-size task as seriously compromised, but report them as one could argue that the single-task conditions still provide a baseline for decision time in the absence of having to monitor a precue and to prepare for a possible switch. Be that what it may, the only significant language-group difference in mixing costs is a monolingual advantage in accuracy. The RT and ES yield BFs of comparable magnitude that favour the null hypothesis.

Gapped groups

Eliminating the monolinguals with L2 proficiency of “3” and bilinguals with a “4” leaves samples sizes of 80 and 83, respectively. There were no significant language-group differences, all p 's > .10.

Step-wise regression analyses

As shown in Table 4, 48% of the monolinguals (23% of all participants) indicated that they had no foreign language instruction. Although their L2 proficiency, L2/L1 ratio, and frequency-of-switching can be assigned a value of zero it does not make sense to talk about attributes of their L2. Rather than allowing this subgroup, because of its considerable relative size, to have undue influence the regression analyses were restricted to participants who rated their L2 in the range of 1–7.

Selection of predictors

Many of the measures of bilingualism are correlated with one another and pose collinearity problems. Self-rated L2 proficiency and the L2/L1 ratio are highly correlated, $r = +0.97$. If competition between the two languages underlies bilingual advantages in EF, then this theoretical consideration supports the selection of the L2/L1 ratio. It is also the case that the zero-order correlations between the L2/L1 ratio and the various measures of performance in the cued-switching tasks are slightly stronger than the correlations with L2 proficiency. Thus, the L2/L1 ratio is included as a predictor in the regression analyses that follow. The proportion of time speaking English is highly correlated with the frequency of switching, $r(192) = -0.77$. As we will report a dedicated analysis of frequency-of-switching in a later section, the proportion of time speaking English is included in the present analyses. Age, PED, and Ravens score were included as possible demographic predictors.

Colour-shape task

In a step-wise regression none of the five predictors (L2/L1 ratio, proportion of time speaking English, age, PED, and Ravens) of switching costs entered the model. This was true for switching-cost RT, accuracy, and ES. Likewise, there were no significant predictors in analyses of the three measures of mixing costs.

Letter-size task

The six step-wise regressions (each using the same set of five predictors) all yielded empty models when the probability of entry was set at the standard .05 level.

Animacy-size task

In a step-wise regression none of the five predictors of switching costs entered the model. For the analyses treating the three mixing-cost measures as the outcome variable, the only significant predictor was the proportion of time speaking English as a predictor of mixing-cost accuracy, $\beta = -.311$, $t = 6.04$, $p = .000$. Given the issues described earlier for the mixing-cost measures in the animacy-size task, this isolated effect will not be discussed further.

Summary of step-wise regression results

The L2/L1 ratio was not a significant predictor of any measure of either switching costs or mixing

costs. Thus, using a continuous measure of bilingualism yielded the same results as those obtained by partitioning the speakers into monolingual and bilinguals, namely, no effects of bilingualism on measures of switching ability.

Single-target verbal-fluency results

The top section of Table 7 shows the mean number of correct responses, the mean number of intrusions (instances of the wrong category), and the mean number of repetitions for both bilinguals and monolinguals and for both single categories and single letters. As is often the case the monolinguals showed a small, but significant advantage in number of correct responses. Intrusions and repetitions did not occur often and there were no differences between the groups with respect to either type of error.

In a 2×2 ANOVA of the two tasks (letters versus categories) and the two groups, the critical Task × Language Group interaction was not significant, $F(1, 203) = .117$, partial $\eta^2 = .001$, $p = .733$; indicating that the monolingual advantage was the same in both tasks and that the bilinguals were not catching up in the letter task. Consistent with the t -test reported above the main effect of Language Group was significant, $F(1, 203) = 6.059$, partial $\eta^2 = .029$, $p = .000$ and in the direction of a monolingual advantage. Somewhat fortuitously the main effect of task was not significant, $F(1, 203) = 1.117$, partial $\eta^2 = .005$, $p = .292$ with the mean number of correct responses across the two target categories equalling 11.6 compared to 11.8 across the two target letters. Thus, the theory-driven prediction of an interaction between Task and Language Group is not complicated by differences in task difficulty, as indexed by mean number of correct responses.

Table 7. Mean number of correct responses, intrusions, and repetitions in each of the verbal-fluency tasks with the standard error of the mean in parentheses.

Task/Group	Correct	Intrusions	Repetitions
Single category			
Bilingual	11.1 (.35)	.24 (.05)	.14 (.02)
Monolingual	12.0 (.35)	.20 (.03)	.18 (.04)
Single letter			
Bilingual	11.3 (.33)	.18 (.04)	.14 (.03)
Monolingual	12.5 (.42)	.22 (.05)	.18 (.04)
Alternating category			
Bilingual	12.3 (.33)	1.13 (.13)	.15 (.03)
Monolingual	13.3 (.32)	1.31 (.14)	.17 (.04)
Alternating letter			
Bilingual	13.4 (.36)	.04 (.01)	.12 (.03)
Monolingual	15.2 (.43)	.05 (.02)	.14 (.04)

Alternating-response verbal-fluency results

The bottom portion of Table 7 shows the mean number of correct responses, the number of intrusions, and the number repetitions for the verbal-fluency tasks that required alternation between two designated letters or two designated categories. These direct comparisons show the same pattern observed for single targets—that is: significant monolingual advantages for the mean number of correct responses but no significant differences between the language groups with respect to the two types of errors.

The right panel of Figure 2 shows the Task × Language Group interaction when participants must alternate between two targets. Despite appearances, the ANOVA showed that the interaction was not significant, $F(1, 202) = 2.144$, partial $\eta^2 = .011$, $p = .145$. It is important to note that the non-significant trends run against the hypothesis that letter fluency is more dependent on EF and that bilinguals should start to catch up with letter targets compared to semantic-category targets.

Verbal-fluency results with matched groups

As reported above the bilinguals had significantly smaller MINT scores. The Group × Task interaction was re-examined for a subset of 59 bilinguals and 59 monolinguals that were precisely matched on their respective MINT scores. Figure 3 shows the distribution of MINT scores for the two original groups. By visual inspection it is apparent that the matching process primarily eliminated the lowest-scoring tail of the bilingual distribution in this test of English vocabulary.

There is a risk that the selected bilinguals differ from those not selected on dimensions that were not measured. Be that what it may; Figure 4 shows the Group × Task interaction for these matched groups. Neither of the main effects nor the Group × Task interaction was significant. Thus, matching on English vocabulary completely eliminated the bilingual disadvantage in verbal fluency, but did not generate a bilingual advantage in the presumably EF demanding letter task. It is worth noting that the trend in the letter condition actually favours the monolinguals.

The right panel of Figure 4 shows the Group × Task interaction for the matched groups performing the alternating-letters and alternating-categories

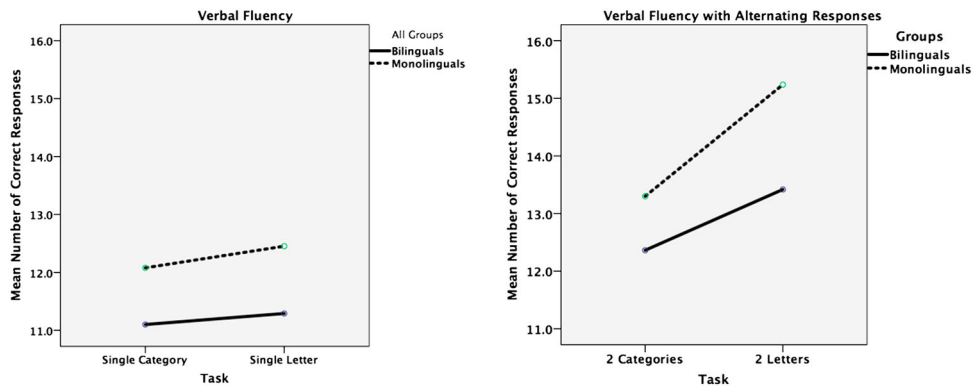


Figure 2. Mean number of correct responses in category versus letter fluency for monolinguals and bilinguals. The results for single targets are on the left and for alternating targets on the right.

tasks. Neither the main effect of group, nor the Group $\times$ Task interaction was significant, both p 's $> .15$. In contrast to the single-target conditions shown on the left, more correct responses were generated in the alternating conditions when the targets were two letters rather than two categories, $F(1, 116) = 21.27$, partial $\eta^2 = .16$, $p = .000$.

Convergent validity between measures of switching costs

Table 8 shows the inter-task correlations between switching costs in the five different tasks: colour-shape, letter-digit, animacy-size, alternating categories, and alternating letters. Switching costs for the three cued-switching tasks significantly correlate with one another with r 's ranging from .250 to .337. This is consistent with the earlier work by Friedman and Miyake showing that these tasks load on the switching factor. The correlation between alternating

letters and alternating categories is also strong (viz., $r = .343$). The correlations between the three cued-switching tasks and the verbal-fluency tasks requiring alternating responses are uniformly smaller and only the correlation between animacy-size and alternating letters is statistically significant, $r = -.148$. In summary, there is reasonable support that the switching costs obtained in the three cued-switching tasks are tapping into a common ability and, similarly, that alternating between letter targets or semantic-category targets in verbal-fluency tasks tap into a common ability. The evidence that this is a "domain-free" switching ability is weak at best.

Frequency of language switching

A composite database was formed by combining the colour-shape switching data obtained in the present study with that used by Paap and Greenberg (2013, Study 3) and by Paap and Sawi (2014). The colour-shape task was identical in all three studies and recruited from the same participant pool with the same recruiting methods and incentives. All three studies included this item:

Some bilinguals switch from one language to the other many times every day because they converse with many others who speak the same languages. Others switch rarely because they only speak

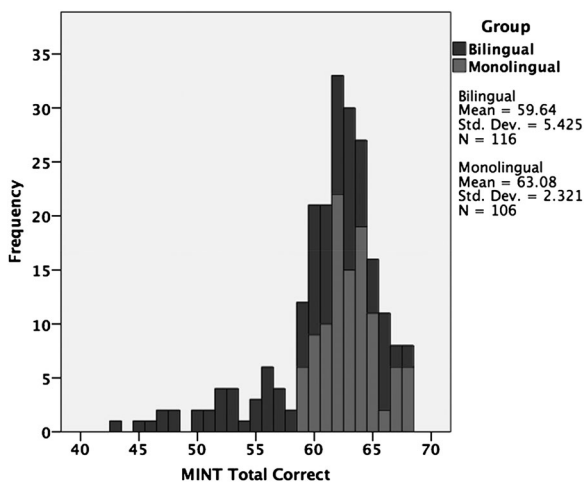


Figure 3. Frequency distribution of MINT English vocabulary scores for bilinguals and monolinguals.

Table 8. Correlations between the five measures of switching costs.

	2	3	4	5
1. Colour-Shape	.250**	.272**	.007	-.021
2. Letter-Digit		.337**	.030	-.003
3. Animacy-Size			-.050	-.148*
4. Alternating categories				.343**
5. Alternating letters				

* $p < .05$, ** $p < .01$.

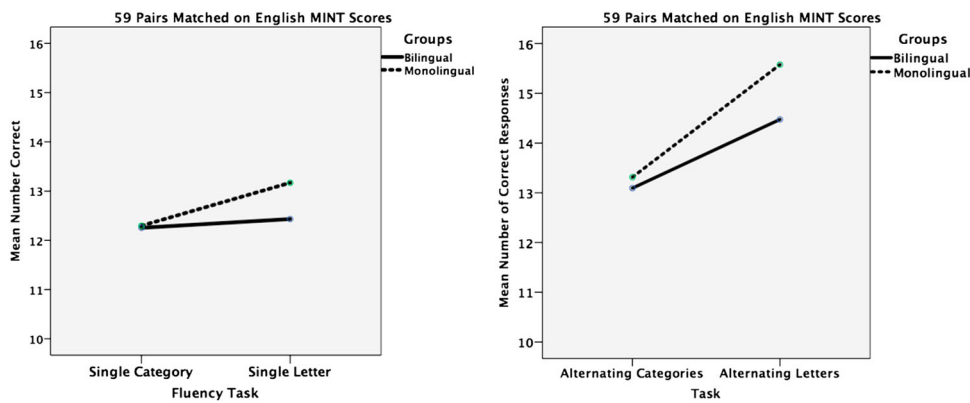


Figure 4. Mean number of correct responses to category and letter targets for the bilingual and monolinguals groups matched on MINT vocabulary scores. The standard single-target condition is on the left and the alternating-targets condition is on the right.

English here at SFSU and only speak their native language when they are at home. I usually switch from one language to the other:

- (1) a couple of times a month
- (2) a couple of times a week
- (3) a couple of times a day
- (4) several times a day
- (5) dozens of times a day

The corresponding item used by Verreyt et al. (2016) was translated from Dutch as: “There, the bilinguals had to indicate how often they switched between languages on a scale ranging from 0 (=never) to 7 (=very often).”⁷ Balanced bilinguals with a rating of 4 or higher were assigned by Verreyt et al. to the high-switch group and those with a rating of 0 to 2 to the low-switch group (no participant rated him/herself as a 3). Although we will also report non-parametric correlational analyses that simply use the five ordinal values, we first approximated the Verreyt et al. bilingual groups by combining bilinguals who responded 1, 2, or 3 into a “low-switch” group and those who responded 4 or 5 into a “high-switch” group.

The mean switching costs (RTs) and 95% CIs for monolinguals ($n = 273$), low-switch bilinguals ($n = 170$), and high-switch bilinguals ($n = 113$) are shown on the left side of Figure 5. There is substantial overlap in the CIs for the three means and one-way ANOVA confirms no significant differences

between the groups, $F(2, 555) = 1.134$, $p = .323$. The mean mixing costs and 95% CIs are shown on the right side of Figure 5. There is a significant effect of groups, $F(2, 553) = 4.556$, $p = .011$. Two orthogonal contrasts were used to follow up the main effect. The mean mixing costs for the monolinguals is significantly smaller than the mean of the two bilinguals groups combined, $t(553) = 2.85$, $p = .005$. However, the low-switch and high-switch groups do not statistically differ, $t(553) = 0.57$, $p = .567$. The monolingual advantage in mixing costs runs counter to the bilingual advantage hypothesis and the absence of any difference between low- and high-switch bilinguals is inconsistent with the hypothesis that FoLS is a key experience in driving bilingual advantages in EF. Taken at face value this monolingual advantage in mixing costs signals the possibility that monolinguals are actually better at monitoring and updating in the colour–shape switching task. However, there is no apparent reason for such an advantage and it does not appear anywhere in the long list of mixing-cost results shown in Table 1.

In a final attempt to uncover any association between frequency of switching (ratings of 1, 2, 3, 4, or 5) and EF we computed Spearman’s rho for both switching costs and mixing costs on the composite dataset for all participants with a self-rated L2 proficiency of 3 or more. The rho values were small, $\rho(320) = -.098$, $p = .080$ and $\rho(320) = -.007$, $p = .904$, for switching costs and mixing costs, respectively.

⁷In the appendix the item is translated as: How often are you in a situation in which you switch between languages? 0 (never) to 7 (very often). One of the authors, EW, clarified that the “how often do you switch” (rather than the “how often are you in a situation”) phrasing is the better translation. In either form the Verreyt et al.’s scale appears to be more open to differences in interpretation as each participant must decide what “very often” means.

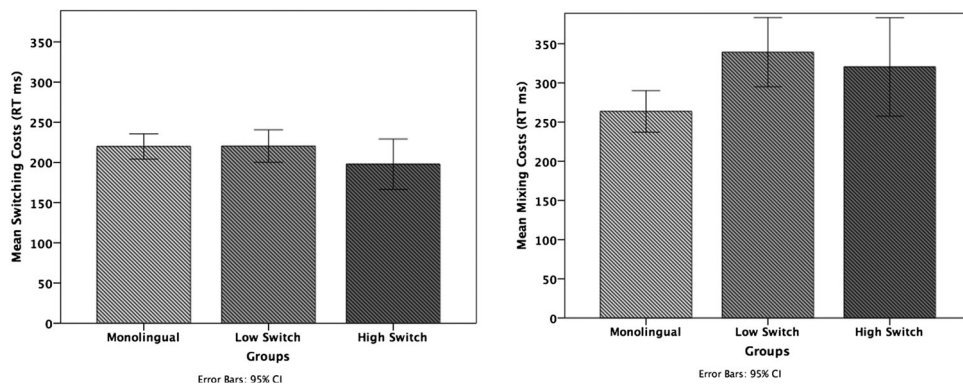


Figure 5. Mean switching costs (left) and mixing costs (right) for monolinguals, low-switch bilinguals, and high-switch bilinguals based on the partitioning of a composite database formed from the current study and two previous studies.

Although the negative ρ for switching costs is in the direction consistent with the hypothesis that switching costs will be smaller with higher rates of language switching, the trending p value is influenced by the large number of degrees of freedom. The FoLS ratings predict less than 1% of the variance in switching costs. Visual inspection of the scatterplot for switching costs shown in Figure 6 supports the general conclusion that in these data FoLS is not a key determiner of switching costs.

Discussion

Bilingual advantages in the switching component of EF

The seminal study by Prior and MacWhinney (2010) reported a substantial bilingual advantage in switching costs obtained with a colour–shape switching task. The null results obtained in the present

colour–shape task sustain the litany of failures to replicate shown in Table 1, especially when participants are young adults. The present results were based on an unusually large sample size and were the same for switching costs based on latency, accuracy, and efficiency. Furthermore, the null results were replicated in two additional analyses; one that eliminated participants with ambiguous proficiencies of “3” or “4” and a second that treated the L2/L1 ratio as a continuous measure of bilingualism. All of these tests were repeated on switching costs derived from the letter–digit and animacy–size tasks and these results were also all null. General switching ability was also evaluated using the alternating-targets variant of the verbal-fluency task, which is part of the popular K-FES battery. These tests resulted in significant monolingual advantages, of comparable magnitude, for both alternating letters and alternating targets. The full-spectrum of results suggests that the bilingual advantages obtained by Prior and MacWhinney (2010) and Prior and Gollan’s (2011) Spanish–English bilinguals may have been false positives.

Additional research using elderly participants would be worthwhile, especially with large sample sizes, multiple switching tasks, and well-matched groups. However, a recent study by Ramos, García, Antón, Casaponsa, and Duñabeitia (2016) is not consistent with the hypothesis that bilingualism improves switching ability in older adults. In this longitudinal study 26 Spanish monolinguals (mean age = 67) received intensive instruction in Basque for eight months. Participants attended class in groups of 10 or less for a total of 5.5 of training per week distributed over three sessions per week. Despite clear evidence that these older adults were learning Basque, there was no improvement in

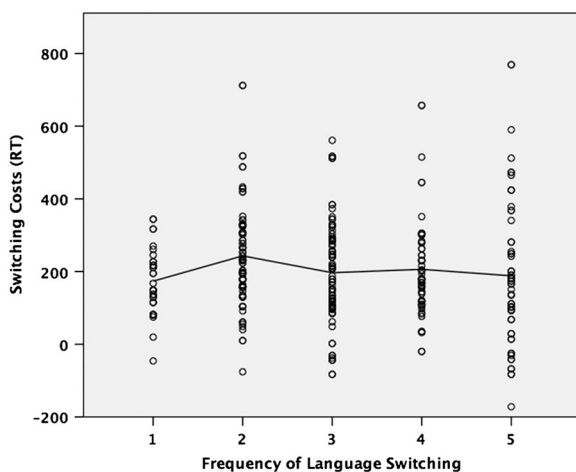


Figure 6. Scatterplot of switching costs at each of the five levels of FoLS.

their switch costs obtained from a colour–shape switching task. A passive control group also showed no improvement and this eliminates the possibility that gains from L2 learning were wiped out by a loss due to aging.

Bilingual advantages in mixing costs

As reviewed in the chronology of results using the colour–shape switching task (Table 1), there has been only one reported bilingual advantage in mixing costs (Wiseheart et al., 2014) using the colour–shape task. In the present colour–shape task the mixing costs based on latency showed a significant monolingual advantage, but that monolingual advantage was not present in either the parallel analysis of ES or in the analysis of gapped groups. Furthermore, there were no significant group differences for mixing costs based on RT or ES in the other two cued-switching tasks (letter–digit and animacy–size). As a monolingual advantage in mixing costs was not observed in any previous study using the colour–shape task and does not generalise to the ES measure or to the other cued-switching tasks used in our study, the monolingual advantage in latency-based mixing costs is a clear anomaly. Consistent with a conclusion that bilingualism does not affect mixing costs, the regression analyses showed that the L2/L1 ratio did not predict mixing costs.

Does FoLS affect switching ability?

The previous and current evidence is consistent with the conclusion that the vast majority of bilinguals do not perform any better than monolinguals in nonverbal switching tasks such as the colour–shape task. As broached in the introduction this should not be the case if there is functional overlap in the mechanisms that involve language switching and general switching.

The study by Verreyt et al. (2016) provided some support for the hypothesis that FoLS is a “key determinant” of bilingual advantages in EF in that high-switch balanced bilinguals had smaller Simon interference effects compared to low-switch balanced bilinguals. This specific difference between low- and high-switch balanced bilinguals did not general-

ise to the flanker task. Furthermore, in their companion study Woumans et al. (2015) reported no differences in Simon or flanker interference between interpreters, balanced bilinguals, and unbalanced bilinguals who probably differed considerably in terms of their FoLS. Paap et al. partitioned the Paap and Sawi (2014) data in a manner very similar to Verreyt et al. (2016) but found no group differences on any measure of EF derived from either the Simon or the flanker task. When the FoLS ratings were directly correlated with the Simon and flanker interference scores there were no significant Spearman correlations.

In summary, there is only one test among many where FoLS is significantly associated with better performance on a measure of inhibitory control. The underlying logic of these tests rests on the assumption that the inhibition and shifting components of general EF are not completely separable and that, indeed, each act of switching languages also leads to the inhibition of the previously active language schema via a domain-free inhibitory-control mechanism.⁸ Given the absence of any systematic relationship between FoLS and measures of inhibitory control, it is worth investigating the more direct link between FoLS and switching costs. When Yim and Bialystok (2012) explored this possibility they found that the correlation between their objective measure of FoLS and a measure of nonverbal switching was statistically zero. Similarly, when the results of the present colour–shape task were combined with two of our earlier studies to form a very large dataset ($n = 320$) the correlations between FoLS and various measures of switching ability were all non-significant despite the large degrees of freedom.

There is no evidence at all that FoLS enhances domain-free switching ability as indexed by switching costs obtained in nonverbal switching tasks. Nonetheless, it may be worthwhile to further investigate the question using more sophisticated measures of FoLS. Yim and Bialystok’s measure was the mean number of switches per minute during a structured interview that yielded an average of 92 seconds of conversation per participant. Although the sample is temporally quite short and the measure may have been influenced by fluency⁹ there was a significant negative

⁸Although the prevailing view is that language switching includes inhibiting the previously active language, that assumption is neither necessary nor universally accepted (Branzi, Martin, Abutalebi, & Costa, 2014).

⁹A fast talker who completes the interview with the same number of switches per words produced would be credited with a higher switch rate.

relationship ($r = -.43, p < .001$) between the switch-rate measure and verbal switching cost derived from a category-fluency task that required participants to alternate between instances from a target category in L1 and L2. This moderate correlation shows that higher frequencies of language switching are associated with lower costs in language switching. However, there were no significant correlations with any of the measures of EF obtained from the two nonverbal tasks. The lack of correlation between the verbal and nonverbal switching results led Yim and Bialystok to conclude that their results "... support a dissociation between the mechanism by which code-switching influences verbal and non-verbal task switching" (p. 881).

Language switching in different interactional contexts

Another limitation of the FoLS construct is that it does not take into account that multilinguals select and switch languages for different reasons and that these motivations vary across different interactional contexts. A key idea behind Green and Abutalebi's (2013) updated Adaptive Control hypothesis is that different interactional contexts trigger different modes of language control. In a pure single-language context, the context-appropriate language is used throughout the interaction. For example, a student's classroom and campus interactions may be conducted exclusively in English, conversations at home in Mandarin, and exchanges at work only in English. The recurrent interactional context for this individual is strongly single-language. In a strong dual-language context, switching is more frequent and triggered primarily by partners entering or leaving the conversation in the same context. Although not explicitly mentioned by Green and Abutalebi more abstract contextual cues such as change of topic presumably induce additional language switches in dual-language contexts. In a dense-code-switching context, not only are there frequent switches, but many occur within sentences including adapting words from one language in the context of the other in an "opportunistic" fashion.

Each context is best served by a different subset of control modes. Dual-language contexts rely on a host of control processes more so than single-language contexts while dense-code-switching does not. Furthermore, if an individual engages

predominantly in one control mode it is assumed that the language-control system will adapt to maximise efficiency, capacity, or connectivity for that mode and context. Green and Abutalebi speculated that adaptation to the dual-language mode is likely to enhance a plethora of control processes: goal maintenance, interference control, salient cue detection, selective response inhibition, and task disengagement and engagement.

Green and Abutalebi (2013) allowed that their predictions may prove false and that there may be no systematic behavioural or neural differences as a function of interactional context. Given that self-scepticism and the impressive rate of citation for the adaptive control paper it is surprising that there have been few direct tests of the model's predictions. The most relevant for present purposes is Hartanto and Yang's (2016) study that compares single-language context multilinguals to dual-language multilinguals in the performance of a colour-shape switching task. The entire sample of multilinguals was divided using a mean split based on ratings of the extent to which they used (1 = never, 5 = always) two languages within the same context or in different contexts. Note that these ratings do not refer to the FoLS per se, but rather the degree to which a single language is used in specific situations. Although multilinguals falling into the dual-language group did have higher rates of intersentential code switching, the average frequency of switching was not high ($M = 2.97, SD = 0.96$ on a 5-point scale). Turning to the results, the dual-language group had significantly smaller switch costs, but there were no group differences on mixing costs. Furthermore in regression analyses context ratings significantly predicted switching costs whereas frequency of intersentential switches was only "moderately" significant. Another informative outcome was that intrasentential code switching positively predicted (i.e. exacerbated) switch costs. If the frequency of intrasentential switches is a valid measure of the degree to which a multilingual operates in an open and "cooperative" control mode, then this too is consistent with the Adaptive Control hypothesis.

The Hartanto and Yang's results are not completely consistent with the Adaptive Control hypothesis. Green and Abutalebi emphasise that in a dual-language context bilinguals must continually monitor for salient cues that might initiate a switch. Thus, it is surprising that there were no differences in mixing costs between the two context groups. An

interesting, but somewhat puzzling result emerges from Hartanto and Yang's reanalysis of their data using Ratcliff's (1978) diffusion model. The model estimated a set of parameters for switch trials versus repeat trials and for both groups separately. There was a dual-language advantage that was manifest in the non-decision-time parameter, not the drift-rate parameter, and this was interpreted as enhancing task-set reconfiguration rather than modulating the amount of proactive interference. There is some risk to this interpretation as the non-decision-time parameter (the one affected by interactional context) includes all the processes before and after the actual decision phase and that includes encoding and response execution as well as the task reconfiguration process required on a switch trial but not on a repeat trial. Furthermore, in the Hartanto and Yang's results the task cue was presented 250 ms prior to the imperative stimulus and any task reconfiguration that took place in that quarter second would dilute differences in the non-decision parameter that were truly due to updating the new task set in working memory. It is important to emphasise that there were no group differences in drift-rate. If a dual-language context adapts the inhibitory processes required to disengage from the previous task set then differences in drift rate would also be expected between the two groups. Another interpretation issue, pointed out by Green and Abutalebi themselves, is circular causation: "Interactional context may shape adaptive response but individual differences (in predispositions and genetic make-up) surely constrain such effect" (p. 522).

In summary the Hartanto and Yang's results offer compelling evidence that interactional context can influence measures of switching ability that cannot be predicted by the frequency of intersentential language switching. On the other hand, it would be fair to say that only some of the differences predicted by the Adaptive Control hypothesis were observed. Unfortunately, the background questionnaire used in the present study does not enable the multilinguals to be partitioned into single-language and dual-language groups, or to identify dense-code switchers, in a manner similar to Hartanto and Yang. Interactional context certainly deserves additional study as Hartanto and Yang's design did not include a baseline group of monolinguals and it has yet to be demonstrated that dual-language multilinguals have smaller switching costs compared to monolinguals.

Verbal fluency

Differences in letter versus category fluency

There were no Language Group $\times$ Verbal-Fluency Task interactions in the present study. This was true both for the standard verbal-fluency tasks where a single target is specified on each trial or in the alternating-targets version. These results are inconsistent with the hypothesis that bilinguals have superior EF and can "catch-up" to the monolinguals on the more EF-demanding letter-fluency task. The present results fall into line with the sizeable majority of studies shown in Table 2 that also reported no Group $\times$ Task interaction. Of course, our bilinguals did not show advantages in mixing cost or switching costs and to the degree that this reflects comparable general EF between the two groups one might actually have been surprised if bilinguals had shown signs of catching up in the letter-frequency task.

Another, not mutually exclusive, interpretation of our results is that the assumption that letter-fluency is more EF demanding than category fluency is incorrect. Shao, Janse, Visser, and Meyer (2014) reported that updating ability (measured by the operation-span task) is strongly correlated with the number of correct responses in verbal-fluency tasks, but that "...there was no evidence that executive control had a stronger effect on performance in the letter than in the category fluency task" (p. 7). If this conclusion is sustained in future research then an important lesson follows, namely, that relatively better performance by a group on letter fluency compared to category fluency cannot be taken as evidence that the group has superior EF. Rather, such a claim must be backed up by independent and direct tests of EF ability.

The bilingual disadvantage in lexical access

Consistent with many prior reports we obtained a bilingual disadvantage in category fluency in both the single-target and alternating-target versions of the task. There are three major accounts of these disadvantages: (1) cross-language interference, (2) weaker links (frequency lag), and (3) smaller vocabularies (within each lexicon). In line with a similar strategy used by Bialystok, Craik, and Luk (2008b) we selected bilinguals with matching vocabulary scores and reanalysed the Group $\times$ Task interaction. If smaller vocabularies were the sole source of bilingual disadvantages in the category-fluency task, then matched bilinguals should perform at about

the same level as monolinguals on the category task. That, of course, is precisely what we observed in both versions of the category-fluency task. On the other hand, if cross-language interference was the sole source, or even a contributing source, of the bilingual disadvantage in lexical access; then presumably the matched bilinguals would suffer from such interference in the category task and still underperform compared to monolinguals. The near identical performance for bilinguals and monolinguals in our matched groups is inconsistent with the assumption that disadvantages accrue from cross-language interference. This issue is somewhat tangential to our main research question regarding the relationship between language switching and general task switching, and accordingly we will simply acknowledge that the literature evaluating the three major accounts for bilingual disadvantages incorporates several additional measures and tasks (Sandoval, Gollan, Ferreira, & Salmon, 2010), and that our findings may supply an important missing piece of the puzzle.

Summary and conclusions

More than 200 SFSU students completed three cued-switching tasks: colour–shape, letter–number, and animacy–size. We replicated Friedman et al.’s (2008) report of significant correlations between the switching costs derived from each task, thus verifying that the three tasks show convergent validity as measures of switching ability. The substantial level of convergent validity felicitously contrasts with the dismal levels of convergent validity for common measures of inhibitory control. This set the stage for a strong and compelling test for bilingual advantages in EF as advocated by Paap and Greenberg (2013) in that multiple measures of the same component of EF (switching in this case) can be derived for both language groups. The results consistently showed no bilingual advantages in switching ability across the tasks and this severely attenuates the chances that they were task specific. These null results contrast with the bilingual advantage in switching costs reported by Prior and MacWhinney (2010), but are completely consistent with the null results reported in most other experiments. Thus, the cumulative evidence favours the conclusion that there may be no differences between bilinguals and monolinguals in switching ability. This is not to say that the pendulum could not swing back. The Hartanto and Yang’s results

suggest that sophisticated measures of interactional context and computational models of the switching task may lead to reproducible bilingual advantages. Similarly, Stocco and Prat’s (2014) exploration of a task using ever-changing rules opens up the forum for a redefinition of switching or flexibility that may be a more productive path for future research.

Returning to our results, participants also responded to two category-fluency probes two verbal-fluency probes, and two probes to alternate between two categories. Consistent with previous research monolinguals generated more correct responses in the category-fluency tasks, but the differences were eliminated when the groups were matched on vocabulary size. Consistent with the majority, but not all of the previous research, the bilinguals did not narrow the gap when targets were letters rather than categories. These findings do not align with the clinical and neuroimaging evidence suggesting that verbal ability may be more strongly reflected in category than in letter-fluency scores, and that, conversely, EF may be more strongly reflected in letter-fluency scores. Of more interest to the relationship between language switching and general task switching, the verbal-fluency measures using the alternating-targets rule generated weak correlations at best with the switching costs obtained in the cued-switching tasks.

Acknowledgements

We thank Angel Avalos, Chirag Dalibar, Jack Darrow, Alina Garrido, Oscar Gonzalez-Munoz, Julie Nacouzi, and Shi-geaki Masuda for their contributions to this project.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Bak, T. H. (2015). Beyond a simple “yes” or “no”. *Cortex*, 73, 332–333.
- Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*, 83(2), 413–422.
- von Bastian, C. C., De Simoni, C., Kane, M., Carruth, N., Coyle, E., Palatella, H., & Miyake, A. (2015, November). *Does being bilingual entail advantages in working memory? A meta-analysis*. Poster presented at the 56th annual meeting of the Psychonomic Society, Chicago, IL.

- von Bastian, C. C., Sousa, A. S., & Gade, M. (2015). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, 141(2), 246–258.
- Bialystok, E., Craik, F. I. M., & Luk, G. (2008a). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 859–873.
- Bialystok, E., Craik, F. I. M., & Luk, G. (2008b). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, 21, 522–538.
- Branzi, F. M., Calabria, M., Gade, M., Fuentes, L. J., & Costa, A. (2016). Beyond the switch cost: Flexible strategies in bilingual minds. Manuscript submitted for publication.
- Branzi, F. M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia*, 52, 102–116.
- de Bruin, A., Bak, T. H., & Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language*, 85, 15–26.
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias. *Psychological Science*, 26(1), 99–107.
- Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81–111). Hove: Psychology Press.
- Calabria, M., Branzi, F. M., Marne, P., Hernández, M., & Costa, A. (2015). Age-related effects over bilingual language control and executive control. *Bilingualism: Language and Cognition*, 18(1), 65–78.
- Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2011). Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology*, 2, 399.
- Christie, J., & Klein, R. (1995). Familiarity and attention: Does what we know affect what we notice? *Memory and Cognition*, 23, 547–550.
- Christoffels, I. K., de Haan, A. M., Steenbergen, L., van den Wildenberg, W. P. M., & Colzato, L. S. (2015). Two is better than one: Bilingual education promotes the flexible mind. *Psychological Research*, 79, 371–379.
- De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain circuit for cognitive control is shared by task and language switching. *Journal of Cognitive Neuroscience*, 27(9), 1752–1765.
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: The Psychological Corporation.
- Donnelly, S., Brooks, P. J., & Homer, B. (2015, July). *Examining the bilingual advantage on conflict resolution tasks: A meta-analysis*. Paper presented at the 37th annual meeting of the Cognitive Science Society, Pasadena, CA.
- Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P., & Tucker-Drob, E. M. (2015). Genes unite executive functions in childhood. *Psychological Science*, 26(8), 1151–1163.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effect of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143–149.
- Festman, J., & Münte, T. F. (2012). Cognitive control in Russian-German bilinguals. *Frontiers in Psychology*, 3, 1–7.
- Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, 6(5), 1–12.
- Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, 137, 201–225.
- Fuller-Thomson, E., & Kuh, D. (2014). The healthy migrant effect may confound the link between bilingualism and delayed onset of Alzheimer's disease. *Cortex*, 52, 128–130.
- García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition, and Neuroscience*, 31(3), 303–327.
- Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience*, 33(2), 387–399.
- Gollan, T. H., Montoya, R. I., & Werner, G. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, 16, 562–576.
- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2011). Self-ratings of spoken language dominance: A multilingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 15(3), 594–615.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5), 515–530.
- 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.
- Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language*, 69, 257–276.
- Hilchey, M. D., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in non-linguistic executive processing tasks. In J. W. Schwieter (Ed.), *Cambridge handbook of bilingual processing* (pp. 586–613). Cambridge: Cambridge University Press.
- Houtzager, N., Lowie, W., & de Bot, C. (2014). Bilingualism, ageing and multilingual processing: Age-related and individual differences in groups of early bilingual Frisians. In L. Filipovitz & M. Pütz (Eds.), *Bilingualism, Cognition and Language use: Processing and typological perspectives* (pp. 263–284). Amsterdam: John Benjamins Publishers.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.
- Johnson, H., Sawi, O., & Paap, K. (2015, March). *Language switching frequency in bilinguals is inconsistently linked*

- to executive functioning. Annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.
- Kirk, N. W., Fiala, L., Scott-Brown, K., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology*, 26(6), 640–648.
- Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology*, 63, 595–619.
- Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience*, 8 (250), 1–12.
- Ljungberg, J. K., Hansson, P., Andrés, P., Josefsson, M., & Nilsson, L.-G. (2013). A longitudinal study of memory advantages in bilinguals. *PLOS ONE*, 8(9), e73029.
- Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, 114(1), 29–41.
- Marian, V., Blumenfeldt, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940–967.
- Marzecová, A., Bukowski, M., Correa, A., Boros, M., Lupiáñez, J., & Wodniecka, Z. (2013). Tracing the bilingual advantage in cognitive control: The role of flexibility in temporal preparation and category switching. *Journal of Cognitive Psychology*, 25(5), 586–604.
- Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin and Review*, 20, 819–858.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychology*, 21(1), 8–14.
- 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, 49–100.
- Mor, B., Yitzhaki-Amsalem, S., & Prior, A. (2014). The joint effect of bilingualism and ADHD on executive functions. *Journal of Attention Disorders*, 19(6), 527–541.
- Moradzadeh, L., Blumenthal, G., & Wiseheart, M. (2014). Musical training, bilingualism, and executive function: A closer look at task switching and dual-task performance. *Cognitive Science*, 39, 1–29.
- Paap, K. R. (2014). The role of componential analysis, categorical hypothesizing, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology*, 26(3), 242–255.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66, 232–258.
- Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, 26(6), 615–639.
- 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.
- Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue? *Cortex*, 74(4), 305–314.
- Paap, K. R., & Liu, Y. (2014). Conflict resolution in sentence processing is the same for bilinguals and monolinguals: The role of confirmation bias in testing for bilingual advantages. *Journal of Neurolinguistics*, 27(1), 50–74.
- Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, divergent validity, and the identification of the theoretical constructs. *Frontiers in Psychology*, 5(962), 1–15.
- Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, 274, 81–93.
- Prior, A., & Gollan, T. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of International Neuropsychological Society*, 17, 1–10.
- Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology*, 25(5), 622–645.
- Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13, 253–262.
- Ramos, S., Garcia, Y. F., Antón, E., Casaponsa, A., & Duñabeitia, J. A. (2016). Does learning a language in the elderly enhance switching ability? *Journal of Neurolinguistics*. Advance online publication. doi:10.1016/j.jneuroling.2016.09.001
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Rodriguez-Pujada, A., Sanjuan, A., Ventura-Campos, N., Román, P., Martín, C., Barceló, F., ... Ávila, C. (2013). Bilinguals use language-control brain areas more than monolinguals to perform non-linguistic switching tasks. *PLoS ONE*, 8(9), e73028.
- Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solis, F. (2000). Verbal fluency and verbal repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, 7, 17–24.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.
- Salthouse, T. A. (2010). Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain and Cognition*, 73, 51–61.
- Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency: The dual-task analogy. *Bilingualism: Language and Cognition*, 13, 231–252.

- Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, 5, 722.
- Shulley, L. J., & Shake, M. C. (2016). Investigating the relationship between bilingualism, cognitive control, and mind wandering. *Journal of Cognitive Psychology*, 28, 257–274.
- Stocco, A., & Prat, C. S. (2014). Bilingualism trains specific brain circuits involved in flexible rule selection and application. *Brain & Language*, 137, 50–61.
- Stoffels, E. J., & van der Molen, M. W. (1988). Immediate arousal and location effects of auditory noise on visual choice reaction time. *Perception & Psychophysics*, 44(1), 7–14.
- Tao, L., Taft, M., & Gollan, T. H. (2015). The bilingual switching advantage: Sometimes related to bilingual proficiency, sometimes not. *Journal of the International Neuropsychological Society*, 21, 531–544.
- Tare, M., & Linck, J. (2011, November). *Exploring bilingual cognitive advantages when controlling for background variables*. Poster presented at the 52nd annual meeting of the Psychonomic Society, Seattle, WA.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York, NY: Cambridge University Press.
- Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, 137, 40–46.
- Verreyt, N., Woumans, E., Vandelandotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, 19(1), 181–190.
- 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, 959–974.
- Wiseheart, M., Viswanathan, M., & Bialystok, E. (2014). Flexibility in task switching by monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 19(1), 141–146.
- Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1579–1586.
- Yim, O., & Bialystok, E. (2012). Degree of conversational code-switching enhances verbal task switching in Cantonese-English bilinguals. *Bilingualism: Language and Cognition*, 15(4), 873–883.