Title
A Relative Bilingual Advantage in Switching With Preparation: Nuanced Explorations of the Proposed Association Between Bilingualism and Task Switching

Permalink
https://escholarship.org/uc/item/2zs4p46f

Journal
Journal of Experimental Psychology General, 146(11)

ISSN
0096-3445

Authors
Stasenko, Alena
Matt, Georg E
Gollan, Tamar H

Publication Date
2017-11-01

DOI
10.1037/xge0000340

Peer reviewed

Journal of Experimental Psychology: General

A Relative Bilingual Advantage in Switching With Preparation: Nuanced Explorations of the Proposed Association Between Bilingualism and Task Switching

Alena Stasenko, Georg E. Matt, Tamar H. Gollan

Funded by:

- National Institute of Child Health and Human Development
- National Institute on Deafness and Other Communication Disorders
- National Science Foundation

A Relative Bilingual Advantage in Switching with Preparation: Nuanced Explorations of the Proposed Association Between Bilingualism and Task Switching

Alena Stasenko[1]

Georg E. Matt [1,2]

Tamar H. Gollan[1,3]

[1]San Diego State University/ University of California, San Diego Joint Doctoral Program in Clinical Psychology

[2]Department of Psychology, San Diego State University, San Diego, CA

[3]Department of Psychiatry, University of California, San Diego

Corresponding Author:
Tamar H. Gollan
Department of Psychiatry, University of California, San Diego
9500 Gilman Dr, La Jolla, CA 92093
Phone: (858) 246-1263
Fax: (858) 246-1287
Email: tgollan@ucsd.edu

Abstract

Bilingual language switching may increase general switching efficiency, but the evidence on this question is mixed. We hypothesized that group differences in switching might be stronger at a long cue-target interval (CTI), which may better tap general switching abilities (Yehene & Meiran, 2007). Eighty Spanish-English bilinguals and 80 monolinguals completed a color-shape switching task, and an analogous language-switching task, varying CTI (short versus long) in both tasks. With longer preparation time (long CTI), bilinguals exhibited significantly smaller task-switching costs than monolinguals, but only in the first half of trials. Group differences diminished with practice, though practice benefitted RTs on short CTI trials more than long, and bilinguals committed fewer errors with practice especially at short CTI. Groups did not differ in mixing costs; however, across CTIs and tasks, bilinguals and monolinguals alike, exhibited robust correlations between mixing costs, but not between switching costs. These results confirm an association between bilingualism and switching efficiency that may be magnified with manipulations that target general switching ability (or could reflect better ability to take advantage of preparation time). However, practice effects observed within experimental paradigms, and between task correlations in costs, may reflect cognitive mechanisms specific to laboratory tasks much more than associations with general switching ability and executive control mechanisms– for which more reliable and valid measures can hopefully be developed in future work.

Keywords: bilingualism, task-switching, bilingual advantage, executive control, color-shape task

A Relative Bilingual Advantage in Switching with Preparation: Nuanced Explorations of the Proposed Association Between Bilingualism and Task Switching

Daily life requires constant flexibility and switching between tasks. Individuals seem to vary in the relative efficiency of their switching abilities, which can have significant consequences for managing and maintaining goal directed behavior, and ultimately professional and personal achievements. A high priority is to identify ways to maximize switching efficiency in any way possible. Implicit in this pursuit might be the assumption that some individuals are better switchers than others. What would characterize performance of an elite switcher? Undoubtedly, an elite switcher would be expected to respond quickly and accurately in a task that measures both switching and non-switching responses, but especially, or perhaps exclusively, on switch trials. We might also expect this switching advantage to be most apparent when individuals are given little time to prepare for upcoming switches. However, this expectation assumes that an integral part of switching ability includes an early stage—prior to execution of the switch itself— in which identification that a switch is needed takes place. Indeed, longer preparation time is known to reduce switching costs in both nonlinguistic (e.g., Monsell, 2003) and linguistic (e.g., Fink & Goldrick, 2015), switching paradigms—but little is known about how switching ability generally relates to preparation time. Finally, an elite switcher advantage might be expected to shrink with practice if elite switchers are more efficient, not because of naturally present switching ability, but because they have had (for one reason or another) more practice with switching than relatively less efficient switchers.

Bilingualism is one reason that might lead some people to switch more often than others. Bilinguals juggle both of their languages on a daily basis, and seem to do so without effort even when switching languages in mid-sentence during conversation with other bilinguals. This juggling process requires managing activation of both languages, possibly relying on the same inhibitory control mechanism thought to facilitate task-switching (Green, 1998; Meuter & Allport, 1999; Myers-Scotton, 2006). In the spotlight of current research is the extent to which mechanisms responsible for selecting which language to speak overlap with cognitive control processes that allow individuals to flexibly switch between non-linguistic tasks (Abutalebi et al., 2012; for reviews see Abutalebi & Green, 2008 Declerck & Philipp, 2015). A main point of focus in this study is whether bilinguals appear to function like elite switchers in a non-linguistic switching task.

A switching advantage for bilinguals would be expected given that bilinguals frequently switch languages, and given a finding in experimental studies of language switching of many striking parallels between results found in linguistic and non-linguistic switching paradigms (e.g., when switching between judging objects by their color versus their shape). In both paradigms, *switching costs* are observed. That is, responses are slower on switch trials (in which the task changes relative to the previous trial), relative to non-switch or *stay* trials (in which the task stays the same as on the previous trial). Also found in both paradigms, are *mixing costs,* which reflect the fact that responses are slower on non-switch trials in a mixed-task block than on non-switch trials in single task blocks (in which just one task is completed). Finally, asymmetric switch costs have been reported in both paradigms such that switching into a more difficult task or less proficient language incurs smaller costs than switching into a less difficult task or more proficient language (for reviews see Kiesel et al., 2010; Declerck & Philipp, 2015). Functional neuroimaging studies also imply shared mechanisms (Abutalebi et al., 2012; Garbin et al., 2010; Abutalebi & Green, 2007, 2008; Ma et al., 2014); overlapping brain regions appear to support linguistic and non-linguistic task switching, most commonly the dorsolateral prefrontal cortex and the anterior cingulate (Branzi, Della Rosa, Canini, Costa, & Abutalebi, 2015; de Baene et al., 2015; Garbin et al., 2010), and there is considerable overlap in brain regions activated specifically on switch trials in both domains (Weissberger, Gollan, Bondi, Clark, & Wierenga, 2015).

Other evidence implies specialized mechanisms for language switching that are not shared with task switching. For example, Weissberger, Wierenga, Bondi, and Gollan (2012) found that aging bilinguals exhibited much greater difficulty with non-linguistic task switching than with language switching. Similarly, Calabria and colleagues (Calabria, Branzi, Marne, Hernández, & Costa, 2015) tested young, middle-aged and older Catalan-Spanish bilinguals and found an asymmetry in non-linguistic switch costs but not in linguistic switching for all age groups. They also found bigger switch costs in older than in young bilinguals in the non-linguistic task, but not in the linguistic task, and switching costs were not correlated across domains. These findings suggest minimal sharing between language control and executive control mechanisms (see also Gollan, Kleinman, & Wierenga, 2015; Gollan, Sandoval, & Salmon, 2011; Prior & Gollan, 2013). Indeed, Gollan et al., (2014) suggested that even when similar patterns are found, different underlying cognitive mechanisms might be involved. For example, switching costs are reduced in both domains when switches are voluntary instead of cued. However, only in the language task did the voluntary advantage appear to be driven by association of each stimulus with just one of two possible responses consistently throughout the mixed block (a “bottom up” responding strategy in which lexical accessibility drives switching behavior; Kleinman & Gollan, 2016).

More consistent with the shared mechanisms possibility, are reports that bilinguals sometimes outperform monolinguals on nonlinguistic control tasks (Baum & Titone, 2014;

Bialystok, 2011; Costa & Sebastián-Gallés, 2014; Kroll & Bialystok, 2013; Valian, 2015 but see Duñabeitia et al., 2014; Hilchey, Saint-Aubin, & Klein, 2015). However, this topic is a matter of somewhat heated debate. Paap, Johnson, and Sawi (2016) suggested that the bilingual advantage is not well replicated and highly variable. On the other hand, a recent meta-analysis confirmed the bilingual advantage and reported moderate effect sizes (e.g., $d = .30$ in de Bruin, Treccani, & Della Sala, 2015). Though this gives reason to continue pursuing the question, meta-analyses, by their nature, gloss over small methodological differences across studies that may be critical for revealing when and where differences between groups can be detected. Such details are important not only for understanding the group differences themselves, but also for constraining the underlying theories that explain the behavior more generally in all populations.

Group Differences on the Color-Shape Task

On the possibility of a shared switching mechanism for linguistic and non-linguistic switching, the bulk of evidence accumulated to date comes from reports that bilinguals exhibit smaller switch costs than monolinguals, most often investigated using a switching task developed by Rubin and Meiran (2005) in which participants switch between judging color and shape based on a visually presented cue. Prior & MacWhinney (2010) were the first to report a bilingual switching advantage using this task. In their study, bilinguals exhibited an elite switcher pattern; they responded more quickly than monolinguals on *switch* trials, but not on other trial types, exhibiting smaller switch costs, and the same size mixing costs, as monolinguals. Garbin and colleagues (2010) also reported reduced errors on switch trials for bilinguals relative to monolinguals, but did not observe any group differences in RTs. Other studies also reported smaller switch costs for bilinguals, but did not conform to the elite switcher pattern. For example, in Prior and Gollan (2011) Spanish-English bilinguals responded more slowly than monolinguals, but slowing was larger on stay trials than on switch trials (significantly smaller after controlling for SES; see below). In the same study, Mandarin-English bilinguals did not differ from monolinguals in speed or the size of switch costs, and the bilinguals also reported that they switched languages significantly less often in daily life than did Spanish-English bilinguals. Interestingly, comparing these two bilingual groups in a *language switching* task, Spanish-English bilinguals exhibited significantly smaller switch costs relative to Mandarin-English bilinguals with a similar pattern (i.e., slowing on stay trials and in this case also faster responses on switch trials; see Figure 2 in Prior & Gollan, 2011). These results imply an association between language and task switching, and reveal some consistency in the tendency for bilinguals to respond more slowly on stay trials whether in a linguistic or non-linguistic task. Furthermore, they suggest that between group differences are caused by habitual language switching specifically, so that only bilinguals who switch languages frequently exhibit reduced switch costs.

This pattern of relative slowing for bilinguals on stay trials has been replicated in different parts of the world with different types of bilinguals. Garbin and colleagues (2010) found that early Spanish-Catalan bilinguals exhibited equally slow responses on switch and non-switch trials (no switching costs), whereas Spanish monolinguals responded more quickly on non-switch trials. Of note, Garbin et al., modified the color-shape task in a number of ways from its original form in part to accommodate neuroimaging protocols (e.g., there was no delay between cue and target; overlapping button-response mappings; written, instead of pictorial, task cues). Such modifications, though seemingly small, may in part be responsible for variation in findings obtained across studies. Table 1 summarizes methodological differences between studies that did versus did not find smaller switching costs for bilinguals relative to monolinguals using the color-shape task – a topic we return to in the General Discussion.

Most recently, de Bruin and colleagues (2015) compared older Gaelic-English bilinguals and monolinguals matched for multiple relevant factors (e.g., lifestyle, SES, education, IQ, gender, and age) on executive control tasks, including the color-shape task-switching paradigm. Bilinguals were divided into two groups including active versus inactive bilinguals—active bilinguals used both languages on a daily basis whereas inactive bilinguals mainly used English. Active bilinguals exhibited significantly and considerably smaller switching costs than monolinguals (almost four times smaller; 34 ms vs. 130 ms, respectively), while inactive bilinguals did not differ significantly from active bilinguals or monolinguals. Response times were equally fast overall for the three language groups, and active bilinguals were slightly but not significantly faster than monolinguals on switch trials, and slightly but not significantly slower than monolinguals on stay trials. De Bruin and colleagues also reported that between group differences in switching cost were no longer significant when switch costs were calculated as a proportion of overall speed. However, proportional adjustments for baseline speed might not be warranted when groups do not differ significantly in overall speed, and the finding of significantly reduced switching costs for active but not inactive bilinguals is in line with previous conclusions that some forms of bilingual language use lead to significantly smaller switch costs. However, failures to replicate the bilingual advantage in task switching have also been reported (Hernández, Martin, Barceló, and Costa, 2013; Paap & Greenberg, 2013; Paap et al., 2016; Paap & Sawi, 2014; Mor et al., 2015; Prior & Gollan, 2013).

A General Switch Mechanism?

The proposition that bilingual language use might lead to more efficient task switching rests on the assumption that there is a general switching mechanism. This is a question that itself is open to discussion. Some have suggested that set-shifting does not represent a global trait that can be captured by a single measure (Deák & Wiseheart, 2015). Of course, if there is no general switching mechanism, then frequent language switching could not possibly benefit task switching. Yehene and Meiran (2007) argued that there is a general switching ability but that not all switching tasks measure it equally well. They tested participants on two non-linguistic task switching paradigms, a shape/size judgment task and a vertical/horizontal task and also varied preparation time, or *cue-target interval (CTI)*, which was either short (116ms) or long (1016ms). A previous study showed strong correlations between switching costs at short and long cue-target interval (Friedman & Miyake, 2004), and thus, Yehene and Meiran (2007) aimed to explore whether this association remains across two different task switching paradigms which would support the notion of a general switch. Their criteria for general ability included a) shared variance across the two task switching paradigms, and b) shared variance with psychometric intelligence (which they also measured). Correlation analyses and structural equation modeling indicated that switching cost with a longer preparation time (i.e., a long CTI) as well as mixing cost, met these criteria; 37% and 15% of variance were shared across shape/size and vertical/horizontal tasks for switching and mixing costs, respectively. In contrast, switching cost with little preparation time (short CTI) was found to be paradigm-specific and did not meet their criteria for general switching ability. They concluded that general switching ability is best tapped by a switching cost with long preparation time.

It might seem surprising that switching costs with relatively short preparation time appeared to provide a weaker measure of task switching ability than costs with longer CTI, because greater preparation time might allow individuals with weaker executive control time to compensate for their weaknesses. However, Yehene and Meiran (2007) suggested that switching cost at short CTI may partly reflect the time needed to process the task cue, a task that might not be accomplished by the executive function system (Logan & Bundesen, 2003; Mayr & Kliegl, 2003; Yehene & Meiran, 2007). On this view, switch costs with short CTI are contaminated by cue-processing abilities, in which case it is not surprising that they provide a less reliable measure of switching ability[1]. Interestingly, to date the majority of studies that examined switching ability in bilinguals relative to monolinguals used a relatively short CTI (ranging from 0 ms to 250 ms; see Table 1) with the exception of a recent study that used a longer CTI (650 ms) that found a bilingual advantage in switching cost for elderly bilinguals (Houtzager et al., 2015). If Yehene and Meiran are correct that switch costs with short CTIs provide a poor measure of general switching ability, then it is perhaps not surprising that it has been difficult to replicate the bilingual advantage in task switching (which in turn should be more robust with a longer cue-target interval).

---

[1] Note that switch costs at short CTI are *not* completely unreliable, they just appear to be less reliable than mixing costs. Switching costs exhibited convergent validity even at relatively short CTIs in studies that examined factor loadings on a general shifting factor (with CTIs ranging from 0 to 150 ms; Friedman & Miyake, 2004; Klauer, Schmitz, Teige-Mocigemba, and Voss, 2010; Miyake & Friedman, 2012). Similarly, moderate correlations are found between different switching tasks (*r*s of .35 to .37; Friedman & Miyake, 2004). In studies with hundreds of participants (orders of magnitude more than typically tested in studies of the bilingual advantage) switch costs also exhibited good test-retest reliability even with relatively short CTI (Paap & Sawi, 2016), though note that in this latter study mixing costs were considerably more reliable than switching (as in Yehene & Meiran, 2007 and see below).

The Current Study

In this study, we tested the hypothesis that between group differences in task switching will be more apparent with a long than with a short CTI. We tested the same population of Spanish-English bilinguals that we previously reported exhibited smaller switching costs than monolinguals using the color-shape switching task (Prior & Gollan, 2011), and that was investigated in follow up studies (Prior & Gollan, 2013; Weissberger et al., 2012; Weissberger et al., 2015). We used a similar task and experimental design, but with a few modifications aimed at achieving a more robust signal. First, following the logic laid out by Yehene & Meiran (2007), we manipulated CTI including the same short (116 ms) or long (1016 ms) intervals used in that study. We hypothesized that if switch costs observed at long CTI offer a better reflection of general switching ability than at short CTI, and if bilingual language switching shares processes with switching in general, then differences between bilinguals and monolinguals should also be more robust at long CTI.

Also following Yehene and Meiran (2007), we included 320 mixed block trials, which is nearly twice as many trials as in Prior and MacWhinney (2010) and Prior & Gollan (2011; 2013). On the one hand, the increased number of trials might increase power for detecting between-group differences by creating a more reliable measure of switch costs. On the other hand, increasing the number of trials could reduce the advantage if, as outlined above, the mechanism of that advantage is transfer of practice from language switching. To address the latter possibility, we conducted separate analyses of the first and second half of experimental trials. Previous studies that examined practice effects in this way revealed inconsistent patterns; in one study monolinguals benefitted more from practicing a conflict-monitoring task than bilinguals, resulting in a bilingual advantage in the first block only (Costa, Hernández, Costa-

Faidella, & Sebastián-Gallés, 2009). Conversely, Abutalebi and colleagues (2011) found a reduced conflict effect with practice for bilinguals in a flanker task in the second of two testing sessions, whereas monolinguals did not show a reduction in conflict effects across testing sessions. In the present study, if between group differences in the magnitude of switching costs are found in the first half but not in the second half, this would be consistent with the hypothesis of transfer of practice from habitual language switching to nonlinguistic switching, that disappears when monolinguals accumulate extensive task-specific practice.

Finally, in addition to between group differences, we also looked within each language group at correlations between task switching and mixing costs at the two CTIs, and two standardized measures of task switching—the Trail Making Test and the Color-Word Interference Test (CWIT) to explore individual differences in switching ability as assessed by different measures. We also tested participants in a language switching protocol to more directly examine the hypothesis of shared control mechanisms (as in Prior & Gollan, 2011; 2013). For these analyses, significant correlations between costs across linguistic and nonlinguistic tasks would support the notion of a general switch and shared processing mechanisms across domains. To maximize comparability of linguistic and nonlinguistic switching, we restricted the number of stimuli to four digits (in previous studies we had nine), a manipulation that also made it possible to test monolinguals in our language switching paradigm (because monolinguals tested herein had enough exposure to a foreign language to name four digits). If only bilinguals exhibited cross-task correlations in switching costs, this would suggest that shared mechanisms for task and language control emerge only with extensive experience using two languages. Conversely, if significant correlations emerged even in monolinguals (who have little to no experience with language control), this would suggest that such effects are driven more by the nature of the tasks themselves than by patterns of language use and experience.

Methods

Participants

Eighty monolingual English speakers (76 right-handed) and eighty Spanish-English bilinguals (73 right-handed) who were undergraduates at the University of California, San Diego (UCSD) participated in the study for course-credit. A power analysis (G-Power; Faul et al., 2007) showed that the number of participants needed to achieve a small to medium effect size (as in previous literature) for a within-between interaction (for the factors of Trial Type and Group; 2 levels in each factor) using a repeated measures ANOVA with $f = .24$, power of $1\text{-}\beta = .90$, and a two-tailed $\alpha = .01$, with an average correlation of .7 among repeated factors, is *22 per group*. In addition, when deciding on sample size, with 80 participants per group we doubled the sample size of previous studies that included approximately 40 subjects (or less) per group (see Table 1), and had similar numbers of participants in each group as Yehene and Meiran, 2007 (n=98).

In the color-shape switching task, two participants (one bilingual, one monolingual) were excluded for having accuracy rates lower than 90% (which was 3.5 standard deviations below the mean for all participants). In the language-switching task, one bilingual was excluded for having an accuracy level lower than 93%, approximately 3 standard deviations lower than the mean for all bilinguals. Participants gave informed consent, and all study procedures were in accordance with the policies of the UCSD Institutional Review Board (IRB). Table 2 shows the characteristics of the participants’ self-reported language history as well as demographic information. Bilinguals in the present study reported switching language often daily, and using Spanish often in their daily lives. San Diego is just 13 miles from the Mexican border, and opportunities to mix languages are plentiful. Many of the monolinguals tested herein had some formal training in Spanish or a different language (see Table 2), but primarily used English only throughout their lives. Thus, the monolinguals were not profoundly and purposefully purely monolingual—they had some exposure to Spanish or other languages either in school (see Table 2 for years of formal language training), or in the environment (e.g., many street names in San Diego are in Spanish), but they were clearly different from the bilinguals in their patterns of language use.

An important difference between bilinguals and monolinguals was that they were not matched on parent education level, often considered as a proxy for socioeconomic status (SES). Prior and Gollan (2011) hypothesized Spanish-English bilinguals' performance on the task-switching measure, though improved by their bilingualism, was also negatively affected by their lower SES relative to monolinguals. Supporting this view, bilinguals in that study exhibited significantly smaller switch costs than monolinguals only after matching for parent education level, and when controlling for response slowing by calculating proportionally adjusted switch costs relative to baseline (similar conclusions were reached using an ANCOVA with parent education level entered to control for SES and including all participants tested). Below we consider the possible role of parent education level in modulating group differences.

Materials and procedure. Participants completed one session of ~2 hours of testing consisting of cognitive and linguistic measures. Computerized tasks were presented using PsychoPy v1.83 (Peirce, 2008) on a Macintosh computer with a 20-inch color monitor. Response times for the non-linguistic task were collected with a response box. Naming times were recorded using headset microphones connected to a response box and were also recorded with a digital recorder. Participants were seated ~60 cm from the monitor. As commonly done in studies of individual differences (see Yehene & Meiran, 2007), all tasks were presented in the same testing order for all participants as follows: language history questionnaire, non-linguistic (color-shape) switching, linguistic (digit-naming) switching, Color-Word Interference Test (CWIT), Trail Making Test (TMT), and Multilingual Naming Test (MINT; in English only for monolinguals, in English then in Spanish for bilinguals). In the language-switching task, monolinguals were instructed to use whatever non-English language they could to name the four digits presented in the current study (for 66 monolinguals this was Spanish). We do not report these data in detail because our primary aim was to investigate between-group comparisons in the color-shape task, and because monolinguals would have had no real prior experience with language-switching, and without proficiency in a second language their performance would reflect different processing mechanisms not of interest here.

Color-shape task. This task was based on the design of the Shape/Size paradigm from Yehene and Meiran, 2007 (e.g., overlapping response buttons, randomized structure, number of trials in each block, cue-target interval) but some elements were modified to match the color-shape task from Prior & Gollan, 2011 and Rubin & Meiran, 2005; e.g., stimuli, ‘sandwich’ design). Participants made color and shape judgments on visually presented stimuli, using button presses to indicate their selection. The target stimuli were either red or green circles (3 cm radius) and triangles (3 cm base, 3 cm height). The task cue for the color task was a color gradient and the cue for the shape task was a row of small black shapes (7 cm by 2 cm). Cues were presented 3 cm above the position where the target stimulus would have been presented. Half of the participants were assigned to a response key combination in which “circle” and “green” responses were mapped to the right button box key, and “triangle” and “red” responses were mapped to the left button box key. The other half of the participants were assigned to a reversed response button combination in which “circle” and “red” were mapped to the right response button, while “triangle” and “green” were mapped to the left response button. To minimize working memory requirements, the response cues were displayed on the bottom left and right corners of the screen (e.g., instead of placement of stickers or a template onto the response box as done in Yehene & Meiran, 2007 and Prior & Gollan, 2011; see Weissberger et al., 2015).

Our use of overlapping button-response mappings is unlike most previous studies (see Table 1). With overlapping button-response mappings just two buttons are used and both tasks are represented by each button (e.g., one button is used for both “red” and “circle” responses, and the other button is used for “green” and “triangle” responses). This contrasts with non-overlapping button-response mappings, in which four different response buttons, and four different fingers are used to respond, with each different response mapped onto a separate button-finger combination. Interestingly, as summarized in Table 1, the few studies with overlapping mappings also found a bilingual advantage of some kind, either in mixing (Gold et al., 2013[2]: Experiments 1 & 2; Wiseheart, Viswanathan, & Bialystok, 2014) or in switching (Garbin et al., 2010 and the current study; see Hartanto et al., 2016 for review). Just one study with overlapping mappings did not find a bilingual advantage in behavioral data, but the color-shape task was modified substantially to accommodate neuroimaging (Rodríguez-Pujadas et al., 2013; they cued response rule initially with a written cue, and then subsequent trials were cued with “switch” or “repeat” instructions). Use of overlapping mappings may turn out to be important because it is arguably more similar to what bilinguals experience during language control and switching (i.e., bilinguals speak both of their languages out of one and the same the

---

[2] Note that some have interpreted these data as mixing costs (e.g., Hartanto & Yang, 2016) while others have interpreted this as a global switch cost (Wiseheart, Viswanathan, & Bialystok, 2014).

mouth). Overlapping mappings also increase task difficulty, via increased response competition (Gade and Koch, 2007), or demands on task-set reconfiguration (Meiran, Chorev, Sapir, 2000; Hartanto et al., 2016), which may increase the chances of observing group differences of various kinds (Hartanto et al., 2016; Wiseheart et al., 2014; Mayr, 2001).

Additionally, we used only one cue per task in previous studies that revealed reduced switch costs for bilinguals on the color-shape task. With just one cue per task, the effects of task-switching versus cue-switching are confounded (Logan & Bundesen, 2003), whereas use of two cues per task would make it possible to tease apart cue processing costs from task switching costs. However, additional cues can increase the working memory load, and therefore we avoided this modification in the interest of replicating previously observed differences between groups, and the possibility that working memory load might also modulate these differences.

Tasks were administered using a sandwich design (Rubin & Meiran, 2005) with one single task block, either color or shape, counterbalanced between subjects (80 trials), followed by four mixed task blocks, which had both color and shape decisions (80 trials, each), followed by one single task block, color or shape decisions (80 trials; whichever block was not completed initially). The single task blocks were counterbalanced such that half of the participants completed the color single block first, and the other half completed the shape single block first. Participants in Yehene and Meiran (2007) completed only one single task block following completion of four mixed task blocks. However, to equate power in switching and mixing cost analyses we had each participant complete single task blocks for both tasks, one before and one after the mixed blocks. Before the start of the mixed task blocks, participants were informed about the transition to the mixed task blocks in which they would perform two tasks, and then they completed 10 mixed task trials serving as practice. This counterbalancing procedure meant that tasks were practiced differently between subjects such that, for example, participants who completed the color task first may show smaller mixing costs for color than people who did not receive any practice with the color single block before completing the mixed blocks. However, we collapsed across individual tasks in our presentation of the results (see below), and an equal number of participants completed color first and shape first (counterbalanced) such that on average these differential practice effects should not be apparent in the analyses presented.

Participants were asked to respond as quickly and accurately as possible. Each trial began after a response in the previous trial and consisted of 1) a fixation crossed presented for 2032 ms, 2) the presentation of the task cue for either 116 or 1016 ms (which constituted our cue-target interval, or CTI manipulation) and remained on the screen until the target appeared and 3) the presentation of the cue and the target stimulus until the response was given, or for a maximum duration of 4 seconds, after which the fixation for the next trial began. The design of the mixed task blocks included 16 critical trial-types (with 2 task cues, 2 stimulus shapes, 2 colors of the stimulus, and 2 CTIs equally represented), which were randomly presented without replacement (within the 16 trials), and then repeated 5 times. All possible 16 trial types were presented at least once before another 16-trial ‘miniblock’ repeated. Additionally, the number of switch and stay trials was not constrained in order to follow the randomized nature of the design of Yehene & Meiran, 2007 (note that in Prior & Gollan, 2011 the number of stay and switch trials was equated by creation of fixed-lists). The fully-randomized presentation of trial types resulted in a switch rate of 53% (*SD* = 3%; range = 47-60%), for both monolinguals and bilinguals. Following Yehene and Meiran (2007) and Prior and Gollan (2011), in our primary analyses we collapsed color and shape decisions and thus had 160 single, approximately 160 stay, and 160 switch trials for each participant. As noted above, Prior and Gollan (2011) and Prior & MacWhinney, 2010, had about half as many stay and switch trials (n = 72), therefore, to facilitate comparison across studies, we present data separately for first versus second half of trials. Because of our counterbalancing procedure (some participants completed single color task blocks before and shape after mixed task blocks, and others vice versa) we did not divide single-task blocks into first and second halves, and thus the means for the single task trials reported below are identical for the first and second halves.

Language task. The design of the language task closely mirrored the non-linguistic task; all aspects of the design were the same with the exception of stimuli, task cues, and response type (i.e., voice recording versus button press). The stimuli in all blocks were four numbers (*two, five, eight,* and *ten*, which in Spanish are *dos, cinco, ocho, and diez,* respectively) and participants were asked to name the digit out loud as quickly and accurately as possible based on the language cue. The task cues were the American flag for English and the United Nations flag for Spanish. CTI and target presentation times were identical to those in the color-shape task. Participants first completed one single-language naming block (either in English or Spanish, counterbalanced across participants), followed by 10 trials of mixed task practice trials, 4 mixed language blocks, and one single-language naming block (English or Spanish, counterbalanced). As for the color-shape task, each block contained 80 trials, and the randomized presentation of trial type resulted in a 53% switch rate (*SD* = 3%; range: 47-59%).

Color-Word Interference Test. The Color-Word Interference Test (CWIT; Delis et al., 2001), an extension of the classic Stroop Test (Stroop, 1935) consists of four conditions comprised of 50 items each that are printed on a single page and placed in front of participants. Time in seconds to complete each condition was recorded for each participant. Baseline conditions assess lower-level functions and are composed of Color Naming (condition 1; naming color patches) and Word Reading (condition 2; reading color names printed in black ink). A third condition, Inhibition, assesses individuals' ability to inhibit the automatic tendency to read the word instead of naming the incongruent ink color (condition 3; e.g., say "blue" when the word "green" is printed in blue ink). The fourth condition, Inhibition/Switching, assesses individuals' ability to switch (signaled by a line-drawn box around the stimulus) between naming the incongruent ink color (Naming condition 3) and reading a word (condition 2). In addition to completion times for each of the 4 conditions, Delis and colleagues (2001) provided contrast measures aimed at isolating certain aspects of executive functioning. These included (a) Inhibition Cost (Inhibition minus Color; is meant to reflect an individual's inhibitory control ability while controlling for baseline color naming speed), (b) Inhibition/Switching Cost (Inhibition/Switching minus the sum of Naming and Reading; meant to reflect the ability to both inhibit and switch while controlling for baseline performance in both color naming and word reading), and (c) Switching Cost (Inhibition/Switching minus Inhibition; meant to reflect switching ability while controlling for inhibition ability).

Trail Making Test. The Trail Making Test (TMT: Reitan, 1992) has two conditions and measures individuals' motor sequencing ability and the ability to flexibly switch between number and letter sequencing. Condition A consists of circles on a single 8 ½ x 11 inch page numbered from 1-25. Participants are asked to draw lines with a pencil to connect the numbers in ascending order as quickly and accurately as possible. In condition B, the circles include both numbers (1-13) and letters (A-L) and participants are asked to connect the circles in ascending order, but alternating between numbers and letters (e.g., 1-A-2-B). If the experimenter notices an error, s/he corrects the error by drawing two lines on the incorrectly drawn line and redirecting the participant to fix the error without pausing the stop-watch. Time in seconds to complete each condition was recorded for each participant.

MINT. The Multilingual Naming Test (MINT; Gollan et al., 2012) consists of 68 black-and-white line drawings, presented in increasing order of difficulty. Participants named the line drawings in English. Bilinguals subsequently named the same pictures in Spanish. The total number of pictures correctly named in each language provides a measure of language proficiency. This task was not timed.

Results

Color-Shape Task

Incorrect responses, responses immediately following an error, and responses slower than 3000 ms or faster than 300 ms, were excluded from the analysis (6% and 5% for bilinguals and monolinguals, respectively). All other responses were analyzed using SPSS 23. Table 3 shows means and standard deviations for all trial types for both groups. Figure 1 shows the results for each condition and participant group broken down by first (Figure 1a) versus second (Figure 1b) half of trials and represents only mixed block trials. Error rates in the mixed block of the color-shape task were low (M = 2.6%, SD =0.02; M = 3.1%, SD = 0.02 for bilinguals and monolinguals, respectively, which were not significantly different, p = .144). As such, we report detailed analyses of error rates only where these are critical for interpreting findings of key interest in the RTs. Additionally, Figure 2 shows error rates in each condition and half of trials for each participant group. In the mixed task blocks, responses were faster, and switching costs smaller, for color relative to shape responses, a main effect of Task, $F(1, 158) = 129.78$, MSE = 22,397, $p < .001$, $\eta^2 = .451$, and a significant interaction between Task and Trial Type, $F(1, 158) = 4.08$, MSE = 6,475, $p = .045$, $\eta^2 = .025$. However, there were no significant interactions between task and group, all *F*s < 2.07, *p*s ≥ .15; thus, to maintain consistency with previous studies (Prior & Gollan, 2011, Prior & Gollan, 2013, Yehene & Meiran, 2007), and to simplify the presentation of the results, we collapsed across the task factor.

Color-shape switching costs. To determine whether there were group differences in task switching efficiency, RTs for condition means in the color-shape task were entered into a four-way repeated measures analysis of variance (ANOVA) with language group as a between-subject factor (bilingual, monolingual), and three repeated measures factors including CTI (long, short), Trial Type (stay, switch), and Experimental Half (first, second). Foreshadowing the results briefly, the predicted 3-way interaction between CTI, Group, and Trial Type (switching costs) was not significant overall. However, there was a highly robust four-way interaction between CTI, Trial Type, Half, and Group, suggesting significantly smaller switch costs for bilinguals relative to monolinguals at long, but not short, CTI, but only in the first half of experimental trials. In the second half, there were no significant group differences at either long or short CTI. Below we report all main and interactions effects in order of complexity but refrain from interpretation of effects that are qualified by the four-way interaction.

Participants responded more quickly when the cue-target interval was long, a significant main effect of CTI, *F*(1, 156) = 1293.37, *MSE* =12,759, *p* < .001, η^2 = .892, more quickly when the task repeated versus on switch trials, a significant main effect of Trial Type, *F*(1, 156) = 148.75, *MSE* = 5,285, *p* < .001, η^2 = .488, and more quickly in the second than in the first half, evidencing practice effects, a significant main effect of Half, *F*(1, 156) = 79.86, *MSE* = 18,994 , p < .001, η^2 = .339). There was no main effect of group (*F* < 1).

Group modulated CTI effects, a two-way interaction between CTI and language group, *F*(1, 156) = 5.92, *MSE* =12,759, *p* = .016, η^2 = .037 (but see higher order interaction below).

Follow-up contrasts revealed that both groups exhibited significant CTI effects (both *p*s $\leq$.001), such that response times were slower when participants had less time to prepare (i.e., in short than in long CTI), and looking within both long and short CTI, there were no significant group differences (*p*s $\geq .236$). Participants also benefitted more from practice at short than at long CTI, a significant interaction between CTI and Half, $F(1, 156) = 54.40$, $MSE = 4{,}784$, $p < .001$, $\eta^2 = .259$. Replicating Yehene and Meiran (2007), switch costs were larger when participants had less preparation time, a significant interaction between CTI and Trial Type, $F(1, 156) = 13.31$, $MSE = 5{,}387$, $p < .001$, $\eta^2 = .079$. All other two-way interactions were not significant (*p*s $\geq .126$).

Participants responded more quickly with practice, especially at short CTI and on switch trials, a three-way interaction between CTI, Trial Type and Half, $F(1, 156) = 4.05$, $MSE = 2{,}970$, $p = .046$, $\eta^2 = .025$. All other three-way interactions were not significant (*F*s <1).

To reveal the nature of the observed four-way interaction between CTI, Trial Type, Half, and Group ($F(1, 156) = 8.45$, $MSE = 2{,}970$, $p = .004$, $\eta^2 = .051$), we conducted 2 x 2 x 2 ANOVAs separately for long and short CTI[3]. At long CTI, bilinguals exhibited significantly

---

[3] To further examine the robustness of the four-way interaction, we conducted additional sensitivity analyses. Using Cook's Distance index to examine the impact of cases with undue influence on our findings, we identified eight participants exhibiting unusually large or negative switching costs. After removing these participants and repeating the repeated measures ANOVA, the four-way interaction remained significant. When excluding these 8 participants, the bilingual switching costs at long CTI in the first half was reduced to 10 ms (monolingual group's switching costs were 51 ms; $p = .003$; Cohen's $d = .48$). Finally, to determine if the four-way interaction was robust to multivariate ANOVA assumptions about model residuals (Maxwell & Delaney, 2004), we examined a mixed effects linear model with unstructured error matrix and obtained Bootstrap estimates of the model parameters based on 1,000 replications using Stata 14.1 [StataCorp, 2015]. The four-way interaction remained significant ($z = -2.90$, $p = .004$). In summary, because the bootstrap results are consistent with our original analyses, and because the Cook's Distance analyses did not suggest cases with undue influence, we interpret the four-way interaction to be statistically reliable and not an artifact of outliers or violations of statistical assumptions.

smaller switching costs than monolinguals, but only in the first half of experimental trials, a significant 3-way interaction between Trial Type, Half, and Group, $F(1, 156) = 7.22$, $MSE = 2{,}982$, $p = .008$, $\eta^2 = .044$; in the first half, bilinguals exhibited only an 18 ms switching cost (SD = 109 ms), less than half that of monolinguals, which was 50 ms (SD = 76 ms), and a significant difference ($p = .036$; Cohen's $d = .34$), whereas in the second half, numbers trended non-significantly in the opposite direction (43 vs 28 ms switching costs for bilinguals and monolinguals respectively, $p = .203$). At short CTI, this 3-way interaction was not significant, $p = .214$); bilinguals and monolinguals exhibited similarly sized switching costs; at short CTI in the first half, switching costs for bilinguals and monolinguals were 81 and 72 ms, respectively ($p = .633$), and in the second half, switching costs for bilinguals and monolinguals were 45 and 62 ms, respectively ($p = .276$). Figure 3 shows switching and mixing costs for each group broken down by CTI and by half.

Of note, the bilinguals' reduction in switching costs at long CTI in the first half of trials appeared as if it might be driven by slowed responses on stay trials relative to monolinguals (as in Prior & Gollan, 2011), or alternatively by faster performance on switch trials than expected for bilinguals given their tendency to respond more slowly in other conditions. Follow-up comparisons favored the latter interpretation. To test which trial type was critical for producing between group differences in switch costs in the first half, we conducted additional analyses to isolate stay versus switch responses. In general, in the mixed blocks, bilinguals tended to respond more slowly than monolinguals — this was most consistently apparent in the first half of trials (see Figure 1). Although this slowing in bilinguals was not significant (there was no main effect group), the means tended in that direction in 6 out of 8 group comparisons in the mixed blocks, with one critical exception, which was switch trials with long CTI in the first half. On these trials, bilinguals responded as quickly as monolinguals, and therefore relatively quickly (compared to themselves) given their tendency to respond more slowly in most other comparison points. The second exception was in the second half on stay trials with long CTI, but group differences in response speed in the second half were negligible, especially at long CTI (see Figure 1). Confirming this interpretation, when analyzing only switch responses, a 2-way ANOVA with CTI and Group as factors revealed a significant 2-way interaction, such that bilinguals responded slower than monolinguals on switch trials at short CTI trials, but no group differences on switch trials at long CTI, $F(1, 156) = 9.13$, $MSE = 59{,}757$, $p = .003$, $\eta^2 = .55$. Critically, when analyzing just stay trials, this same 2-way interaction was not significant, ($F < 1$). In these two analyses, matching those reported above, there were no main effects of group (ps $\geq .21$). Thus, it appears that switch trials, not stay trials, drove group differences in switching costs in the first half of the mixed block at long CTI; bilinguals' reduction in switching costs was driven by their relatively faster than expected for them performance on switch trials at long CTI. Additional analyses (presented in Supplemental Materials) illustrated consistency of these apparent patterns across the distribution of RTs (including slowest and fastest responses).

Switch costs in error rates

As noted above, we do not report detailed analyses of errors because participants committed very few errors overall. However, some consideration of errors is critical to ensure that the conclusions drawn were not driven by tradeoffs between speed and accuracy (or any tendencies in this direction). Thus, we repeated our four-way ANOVA, and the critical follow-ups with error rates, and error rates with an arcsine transformation (Winer et al., 1971). Results reported in detail are from untransformed data, and we mention analyses with arcsine transformation only where these produced different results than the untransformed data.

Briefly summarized, the four-way interaction reported above for RTs was just marginally significant in the error data (p = .083; and not significant after arcsine transformation, F<1). However, there was a significant 3-way interaction between Half, CTI, and Group (p = .006) such that bilinguals tended to be more accurate than monolinguals in every condition except in the first half at short CTI where bilinguals and monolinguals had equivalent in error rates. Additional 3-way ANOVAs with Half, CTI, and Trial Type conducted separately for each language group revealed that bilinguals significantly improved their accuracy with practice on short CTI trials, a significant 2-way interaction between Half and CTI (F(1,78) = 4.20, MSE = .002 p = .030, η^2 = .059). This same interaction was not significant in monolinguals (p = .120). Most importantly, as shown in Figure 2, bilinguals tended to commit fewer errors than monolinguals on both stay and switch trials at long CTI (though not significantly so, ps ≥ .187), thus the relative advantage in switching costs could not be attributed to a speed-accuracy trade-off. The only condition in which bilinguals appeared to commit more errors than monolinguals was short CTI switch trials in the first half (though again this was not significant, p = .227), and by the second half bilinguals in fact committed significantly *fewer* errors than monolinguals in this same condition (p = .027; see Figure 2).

Summarizing the results for color-shape switch costs reported thus far, in the first half, bilinguals exhibited significantly smaller switch costs than monolinguals at long but not at short CTI. Similar to previous studies (Prior & Gollan 2011; de Bruin et al., 2015; Garbin et al., 2010) the bilinguals’ reduction in switch costs appeared to reflect slower responses relative to monolinguals on stay trials. However, additional analyses instead implicated switch trials as being relatively faster for bilinguals at long CTI in the first half (when compared with themselves relative to monolinguals on other trial types as the baseline), and therefore switch trials as the critical trial type driving between group differences. Finally, both bilinguals and monolinguals tended to respond more quickly with practice, but only bilinguals also produced fewer errors with practice (at short CTI), whereas monolinguals did not become more accurate with practice.

Color-shape mixing costs. To examine between-group differences in mixing costs, we conducted another four-way repeated measures ANOVA but comparing responses in the single-task blocks with non-switch trials in the mixed blocks in the Trial Type factor (single block, mixed block-stay)[4]. Foreshadowing the results, although mixing costs appeared to provide a more consistent measure of individual differences in task performance than switching costs (a point that will be explained in detail below), we found no significant differences between bilinguals and monolinguals in these analyses (matching results reported by Prior & Gollan, 2011 and Prior & MacWhinney, 2010; see Table 1 for a summary of studies that found mixing cost differences between groups); the four-way interaction that was significant above in the analysis of switching costs was not significant in the analysis of mixing costs (F (1,156) = 1.69, MSE =3,690, p = .196, η^2= .01).

Participants responded more quickly at long than at short CTI, more quickly on single-block than on mixed block trials, and more quickly in the second than in the first half of trials; main effects of CTI, $F(1, 156) = 696.81$, $MSE = 5{,}959$, $p < .001$, $\eta^2 = .817$; Trial Type, $F(1, 156) = 488.52$, $MSE = 59{,}971$, $p < .001$, $\eta^2 = .758$, and Half, $F(1, 156) = 59.32$, $MSE = 5{,}441$, p

---

[4] Note that single block RTs are identical for the first and second experimental halves because of the nature of the counterbalancing procedure (see Methods). However, the contrast between first versus second half nevertheless reveals how mixing costs change with practice.

$< .001$, $\eta^2 = .276$. The main effect of group was not significant, $F(1, 156) = 1.64$, $MSE = 188{,}359$, $p = .202$, $\eta^2 = .010$).

Replicating Yehene and Meiran (2007), participants exhibited larger mixing costs when they had less preparation time, a significant interaction between CTI and Trial Type, $F(1, 156) = 610.02$, $MSE = 5{,}059$, $p < .001$, $\eta^2 = .796$; participants responded more quickly with practice, but more so at short CTI, a significant interaction between CTI and Half, $F(1,156) = 18.36$, $MSE = 2{,}184$, $p < .001$, $\eta^2 = .105$; finally, mixing costs decreased in the second half of the experiment, a significant interaction between Trial Type and Half, $F(1,156) = 59.32$, $MSE = 5{,}441$, $p < .001$, $\eta^2 = .276$. All other two-way interactions were not significant (all ps $\geq .08$).

Mixing costs decreased with practice, especially at short CTI compared to long, a significant three way interaction between CTI, Trial Type and Half, $F(1, 156) = 18.36$, $MSE = 2{,}184$, $p < .001$, $\eta^2 = .105$. All other three-way interactions did not reach significance (ps $\geq .09$).

Summarizing these analyses, there were no between group differences in the magnitude of mixing costs. Looking within each language group separately, the results also appeared to be more similar than different. In monolinguals, mixing costs were significantly reduced in the second half of trials, and more so for short than for long CTI conditions, a significant interaction between CTI, Trial Type, and Half: $F(1, 78) = 19.53$, $MSE = 1{,}744$, $p < .001$, $\eta^2 = .200$. In bilinguals, mixing costs were also reduced in the second half, especially at short CTI; though in this case, the three-way interaction was only marginally significant, $F(1, 78) = 3.71$, $MSE = 2{,}623$, $p = .058$, $\eta^2 = .045$ (see Figure 3).

Language Task

Error rates in the mixed block for bilinguals in the language task were low (*M* = 2.5% *SD* = 0.02). Similar to the color-shape task, we do not report analyses of error rates in detail, but means and *SD*s by condition and half are shown in Table 4. RTs were trimmed using the same exclusion criteria as for the color-shape task (described above), which resulted in exclusion of 5.0% of the data. Figure 4 shows the results for each condition broken down by first (Figure 4a) versus second (Figure 4b) half of trials and contrasts language switching with color-shape switching performance. For reasons explained above, we did not report monolinguals' language switching data in detail. However, briefly summarizing their performance, monolinguals responded much more quickly in English than in whatever other language they used to name the 4 numbers used in the language switching task. Thus, despite massive repetition of just four number names in their non-English language, they, unlike the bilinguals tested in this study, exhibited strong language-dominance effects.

Language switching costs. RTs for condition means were entered into a four-way repeated measures analysis of variance (ANOVA) with Experiment Half (first, second), Language Dominance (Dominant, Non-Dominant), CTI (long, short), and Trial Type (stay, switch), as repeated measures factors. Language dominance was determined by performance on the MINT; bilinguals with higher scores in English than in Spanish were classified as English-dominant (*n* = 73), one bilingual with equivalent English and Spanish scores was also classified as such because of immersion in an English dominant environment, and those with higher Spanish than English scores were classified as Spanish-dominant (*n* = 5). Importantly, the factor in these analyses was dominance, not language, and so this factor had the same number of subjects as all the others. For example, for English-dominant bilinguals, we classified all English responses as dominant and Spanish responses as non-dominant. Conversely, for Spanish-dominant bilinguals, all Spanish responses were classified as dominant and English responses as non-dominant. We collapsed across the language dominance factor in the figure because dominance did not interact significantly with the other conditions, with the exception of one marginally significant interaction that is mentioned briefly below.

Bilinguals exhibited significant switch costs ($M = 54$, $SD = 33$), responded more quickly at long than at short CTI ($M = 643$ vs. $M = 721$), and responded more slowly in the second than in the first half of experimental trials[5] ($M = 664$ vs. $M = 701$); main effects of Trial Type, $F(1, 78) = 200.92$, $MSE = 4{,}367$, $p <.001$, $\eta^2 = .720$; CTI, $F(1, 78) = 406.94$, $MSE = 4{,}781$, $p < .001$, $\eta^2 = .839$, and Half, $F(1, 78) = 45.16$, $MSE = 9{,}711$, $p < .001$, $\eta^2 = .367$. One of these effects marginally varied by language dominance, such that bilinguals responded more quickly at long CTI, but more so for dominant language responses, an interaction between CTI and Dominance, $F(1, 78) = 3.86$, $MSE = 1{,}345$, $p = .053$, $\eta^2 = .047$. All other interactions with language dominance were not significant (Fs ≤ 1.22, ps .27). Bilinguals responded more quickly in the first than in the second half, especially at the long CTI, a significant interaction between Half and CTI, $F\ (1, 78) = 7.14$, $MSE = 2{,}436$, $p = .009$, $\eta^2 = .084$. Notably, unlike the color-shape task, switch costs were not significantly smaller with long than with short preparation time, although the numbers trended in this direction, $F\ (1, 78) = 1.97$, $MSE = 2{,}249$, $p = .164$, $\eta^2 = .025$. No other interactions were significant (ps $\geq .360$). Our finding of faster responses with longer preparation time is consistent with other language switching studies that manipulated preparation time (Costa & Santesteban, 2004; Fink & Goldrick, 2015; Philipp, Gade, and Koch, 2007; but

---

[5] This slowing of responses by the second half of the task could possibly be explained by general fatigue, as the language switching task was completed after the color-shape switching task.

see Verhoef et al., 2009 who found no benefit of preparation on L1 stay trials). However, unlike some previous studies (as in Costa & Santesteban, 2004; Fink & Goldrick), we did not find that switch costs were reduced with longer preparation time.

The finding that responses slowed across experimental halves (with practice) was accompanied by a significant effect in the opposite direction in the error rates. Specifically, at long CTI, bilinguals produced significantly fewer errors with practice; this effect was significant on switch trials (p = .001), but not on stay trials (p = .67). Similarly, at short CTI, bilinguals tended to produce fewer errors in the second than in the first half on switch trials (p = .057), but not on stay trials (p = .110; see Table 4). Analyses with the arcsine transformed data revealed the same effects. Thus, in both linguistic and non-linguistic switching tasks, bilinguals exhibited some significant tendencies to improve their accuracy with practice especially on switch trials. However, this came with a cost in response times only in the language task, but note that this cost was very small; Figure 4 illustrates that in general the language task appeared to largely be at ceiling for bilinguals and changed very little with practice when compared with the color-shape task in which responses sped considerably across experimental halves.

Language mixing costs. Following our analyses of the color-shape task, we also carried out a four-way ANOVA with the same factors as in the switching costs analysis, but replacing switch trials with single task block responses in the Trial Type contrast. Bilinguals exhibited significant mixing costs (M = 116, SE = 9), responded more quickly when the CTI was long (M = 570, SE = 10), and more quickly in the first half of experimental trials; (M = 586, SE = 10); main effects of Trial Type, F (1, 78) = 190.96, MSE = 22,307, p < .001, η^2 = .710; CTI, F (1, 78) = 295.78, MSE = 2,916, p < .001, η^2 = .791, and Half, (F (1, 78) =43.53, MSE = 2,875, p < .001, η^2 = .358. A main effect of language dominance was marginally significant (F (1, 78) = 2.80, *MSE* = 5,887, $p = .099$, $\eta^2 = .035$, such that dominant-language responses were very slightly faster than non-dominant language responses (*M* = 593 vs. 600). Mixing costs generally did not vary by language dominance with the exception of one marginally significant 3-way interaction between dominance, CTI, and Trial Type, $F(1, 78) = 2.96$, $MSE = 1{,}130$, $p = .090$, $\eta^2 = .036$. As in the switching cost analyses, bilinguals responded more quickly in the first than in the second half, especially in the long CTI condition, a significant interaction between Half and CTI, $F(1, 78) = 5.01$, $MSE = 1{,}128$, $p = .028$, $\eta^2 = .060$. Bilinguals also exhibited larger mixing costs in the second part of the experiment, a significant interaction between Half and Trial Type, $F(1, 78) = 43.53$, $MSE = 2{,}875$, $p < .001$, $\eta^2 = .358$. Mixing costs were larger at short than at long CTI, a significant interaction between CTI and Trial Type, $F(1, 78) = 116.09$, $MSE = 1{,}375$, $p < .001$, $\eta^2 = .598$. Finally, there was a larger CTI effect in mixing costs in the first than in the second half, a significant three-way interaction between Half, Trial Type and CTI, $F(1, 78) = 5.01$, $MSE = 1{,}128$, $p = .028$, $\eta^2 = .060$. No other interactions were significant (*p*s ≥ .408).

In summary, like the color-shape task, the language task exhibited significant interactions between CTI and costs, so that generally switching and mixing costs were larger when bilinguals had less preparation time, however this pattern was significant only for mixing costs. Additionally, in the language task bilinguals responded slightly more slowly with practice likely related to an accompanying and significant decrease in error rates with practice on switch trials with both long and short CTIs (whereas in the color-shape task RTs decreased with practice, and the reduction in error rates was found only at short CTI).

Association of Switching Costs Across Domains

The hypothesis of a general switch mechanism that serves both linguistic and non-linguistic switches predicts that switch costs should be correlated across domains. To investigate this possibility, we examined switching and mixing cost correlations between color-shape and language switching tasks, and for comparison we also included two other commonly used tests of switching ability (the Color-Word Interference test, or CWIT, and Trails A and B), and our objective measure of language proficiency (i.e., the MINT). To measure reliability of switch costs within each domain, we also correlated switching and mixing costs across long and short CTIs in each participant group within each task. Tables 5 and 6 show Pearson bivariate correlations between our primary experimental tasks for the bilinguals and monolinguals, respectively[6], and Figure 5 shows scatterplots for the main correlations of interest. After correcting for false discovery rate (FDR) in multiple comparisons using the Benjamini-Hochberg procedure (Williams, Jones, Tukey, 1999; Thissen, Steinberg, & Kuang, 2002) different cut-offs were determined for bilinguals and monolinguals by arranging p-values from lowest to highest and choosing the largest p-value that is smaller than the Benjamini-Hochberg critical value as threshold. We interpreted any effects significant at p of $\leq .009$ and $p \leq .006$, for bilinguals and monolinguals, respectively.

Of greatest interest was the possible relationship between color-shape and language tasks. In both subject groups, with mixing costs as the measure, these cross-domain correlations were moderately sized and statistically significant at both long and short CTI, whereas switch costs were not significantly correlated across domains (see Tables 5 and 6 and Figure 5A-D). This

---

[6] We did not split the data into first and second halves for this analysis in part because the patterns observed were highly similar across halves, and because in the analysis of mixing costs only the stay trials vary with half (see Footnote 4).

same general pattern was found in the within-task correlations (i.e., contrasting long CTI with short CTI) – both bilinguals and monolinguals showed significant positive correlations with mixing costs as the measure, but not with switching costs (see Figure 5E-F). To some extent, standardized clinical measures of switching ability (CWIT) exhibited the same general pattern; significant correlations between mixing costs, but not switching costs, in the color shape task and CWIT measures (see Tables 5 and 6)[7]. Although, the bilingual group displayed many more significant correlations with the CWIT than the monolinguals. Additionally, only the bilinguals showed robust correlations between mixing costs in the language task and the CWIT measures (these correlations were generally absent in monolinguals). In addition, in a previous study we found a bilingual advantage on the CWIT (Tao et al., 2015), but here we found only non-significant trends in this direction (see Table 2). Finally, measures of language proficiency (as measured by the MINT) were not correlated with mixing and switching costs. This too contrasts with findings by Tao and colleagues (2015), who found that higher English MINT scores predicted higher switching costs (as measured by the CWIT), while higher Heritage Language MINT scores predicted lower switching costs. However, caution is warranted in interpreting any between-study differences on these tasks because of prior completion of the color-shape task and the language-switching tasks in the present study.

In summary, mixing costs in the color-shape task revealed multiple robust correlations within- and between tasks (with language switching and with other tests of switching ability in both bilinguals and monolinguals). In contrast, switching costs revealed few such correlations, and even when significant these tended to be smaller than correlations with mixing costs.

---

[7] This might not be entirely unexpected, given that most CWIT measures make no attempt to separate mixing costs from switching costs, with the possible exception of the CWIT Switching Cost – though even here trial to trial performance may differ in important ways from block-wide performance on such tasks.

The Role of SES in Task Switching

As noted above, bilinguals in the present study likely had significantly lower SES than monolinguals as measured by average parental education. An important question that follows is would group differences in switching costs be more robust if controlling for the bilinguals' SES disadvantage? In Prior and Gollan (2011), the bilinguals exhibited significantly smaller switch costs than monolinguals only after controlling for SES. Similarly, in past studies, SES appeared to modulate language group differences in executive functioning, both in children (Calvo & Bialystok, 2014; Carlson & Meltzoff, 2008) and adults (Paap et al., 2015 for review). In addition, studies in monolinguals have linked lower SES to poorer executive function mainly in children (Hackman et al., 2015; Hackman, Farah, Meaney, 2010 for review) but also in middle to late adulthood (Turrell et al., 2002).

In this respect, it is noteworthy that bilinguals in the present study did not respond more slowly than monolinguals (as in Prior & Gollan, 2011). This attenuates possible concerns that the previously reported result may have been spuriously caused by matching confounds (or a misguided approach to statistical analyses, contra suggestions of Paap, Johnson, & Sawi, 2015). The absence of significant response slowing in the present study for bilinguals, despite a clear SES disadvantage could suggest that bilingualism offsets the negative effects of SES on switching ability (Carlson & Meltzoff, 2008; Gollan et al., 2011), or alternatively that the task in the present study was less sensitive to SES effects than previous implementations of the task (e.g., Prior & Gollan, 2011; though the latter possibility seems unlikely given our discussion above concerning overlapping response mappings and task difficulty).

However, somewhat unexpectedly, in the present study monolinguals, but not bilinguals, showed a significant association between parent education level and task-switching costs (see

Tables 5 and 6 and Figure 6). Monolinguals with more educated parents exhibited smaller switching and mixing costs at long CTI ($r = -.24$; $p = .032$ and $r = -.30$, $p = .008$) and smaller mixing costs at the short CTI ($r = -.38$, $p = .001$; but not switching cost, which unexpectedly exhibited a difficult to interpret positive correlation between parent education level and switch costs at short CTI, $r = .32$, $p = .004$). In contrast, and even though bilinguals had parents with much more variable education levels, and far lower levels of education than monolinguals (see Table 2, and black dots in Figure 6), there were no significant correlations in the bilingual group between parent education level and switching or mixing costs at either long CTI ($r = .19$, $p = .09$ and $r = -.08$, $p = .48$, for switching and mixing costs, respectively) or short CTI ($r = .02$, $p = .88$ and $r = -.11$, $p = .33$, for switching and mixing costs, respectively; see Figure 6).

General Discussion

The present study investigated the existence of a general switch mechanism that might be used by bilinguals to control when and how they switch between languages. Critically, following suggestions of Yehene & Meiran (2007) and others (Logan & Bundesen, 2003; Mayr & Kliegl, 2003), we assumed a distinction between cue processing and switch mechanisms, and manipulated preparation time to obtain a purer measure of switching abilities. Specifically, we hypothesized that extending preparation time (by using a longer cue-target interval) would reduce contamination of switch costs by cue processing times (which should not differ between bilinguals and monolinguals), thus revealing more robust between group differences in switching ability (a cognitive mechanism that might be expected to benefit from bilingual language use).

A Shared General Switch Mechanism

As predicted, bilinguals exhibited significantly smaller switch costs than monolinguals when allowed to prepare to switch with the longer CTI, and no analogous between-group difference in switching costs were found with short preparation time. Importantly, the bilinguals' switch cost advantage appeared to be driven by their relatively faster performance on switch trials on which they responded as quickly as monolinguals, whereas in other conditions bilinguals tended to respond slightly but not significantly more slowly than monolinguals. On this view, bilinguals responded faster on switch trials with preparation and prior to extensive practice than expected relative to themselves when compared with monolinguals on other trials in the first half of the mixed-block (a pattern that appeared to be highly consistent across the distribution of RTs; see Supplemental Material).

Though error rates were very low overall in the present study, the hypothesis of a general switch mechanism that may be improved by bilingual language use, was generally consistent with the pattern of results found in the error data. Mostly importantly, there was no evidence that bilinguals committed more errors than monolinguals in the same condition that revealed a bilingual advantage (see Figure 2); i.e., on switch trials at long CTI in the first half, if anything bilinguals tended to be more accurate than monolinguals. Additionally, by the end of the experiment monolinguals increased their speed without any change in accuracy, bilinguals sped up in the responses and either maintained their tendencies towards higher accuracy than monolinguals (at long CTI), or were significantly more accurate than monolinguals (at short CTI in the second half; see Figure 2).

Additionally, bilinguals exhibited reduced switching costs only in the first half of the experiment, which is consistent with the idea that switching ability can improve with practice – a requirement for the hypothesis that bilingual advantages reflect increased practice with switching and transfer from language switching to general switching ability. Previous studies have also reported fleeting bilingual advantages (e.g., Costa et al., 2009) which could imply that whatever transfers from language control to benefit nonlinguistic control (switching in the present study) will pale in comparison with task-specific practice. This is not very surprising given the many differences between how bilinguals use their two languages in daily life and a necessarily limited relationship to how switching is measured in the lab, provided that task-specific switching mechanisms exist and can also benefit from practice. At minimum, the present results suggest that practice effects can modulate between group differences and should be examined (none of the studies shown in Table 1 considered practice effects).

Our interpretation of these data is consistent with previously proposed suggestions of a bilingual advantage in switching (Garbin et al., 2010; Houzager et al., 2015; Prior & MacWhinney, 2010; Prior & Gollan, 2011), and the proposal of a general switch mechanism, and builds on these notions in a number of ways. As noted briefly above, the finding that the bilingual advantage disappears with practice implies that task-specific switch mechanisms exist or develop, and that these domain-specific mechanisms are more powerful than the general switch mechanism and transfer between domains. Furthermore, the observation of a bilingual advantage only at long but not short CTI has additional implications. One view of these findings is that bilinguals are better specifically at switching per se, but they cannot prepare to switch better than monolinguals can (i.e., bilinguals are only better at switching once fully prepared to do so). However, an alternative possibility with a very different underlying cognitive mechanism is that bilinguals switch more efficiently only when given ample time to prepare. On this view, bilinguals are not better at switching per se, but are better able to take advantage of preparation time – meaning they are better than monolinguals at preparing to switch (see Hernández et al., 2013 for possibly related discussion of bilingual advantages in more complex switching tasks).

Yehene & Meiran's (2007) suggestion that switching ability is better measured at long CTI seems more consistent with the first possibility, however either explanation appears to be equally adequate for interpreting the group differences reported herein.

Aspects of the Data that Appear to be Inconsistent with a Shared General Switch Mechanism

Other aspects of our data seem inconsistent with the idea of a shared general switch mechanism. In particular, bilinguals did not respond more quickly than monolinguals on switch trials (for discussion see Wagenmakers, 2015; Hilchey et al., 2015) – though our analysis above implicates switch trials as critical. Additionally, though bilinguals exhibited a disadvantage in socioeconomic status (SES, as measured by parental education level), and though SES appeared to modulate group differences in previous work (Prior & Gollan, 2011), no such pattern emerged with respect to SES in the present study. Switching costs were not robustly correlated with SES in either bilinguals or monolinguals in the present study, and mixing costs were significantly reduced by higher SES only in monolinguals (see Figure 6, Tables 5-6). Possibly critical differences relative to previous studies include our manipulation of CTI and overlapping button-response mappings in the present study. Both of these manipulations arguably make the task more difficult, and overlapping button-response mappings seem critical given our review of the literature in Table 1. Increased task-difficulty might make it easier to observe beneficial effects of bilingualism which might in some circumstances undo or minimize negative effects of very low SES on switching and mixing costs for this group. Importantly, under this interpretation, better matched groups of bilinguals and monolinguals might not necessarily reveal more consistent between group differences (particularly if matched groups removed all participants with very low SES).

Perhaps the greatest challenge to both the general-switch mechanism, and shared-switch, hypotheses, was our failure to observe robust correlations between switching costs across task-switching and language-switching (see Tables 4-5; Figure 5). The total absence of significant correlations here would seem to be at odds with the notion of a general switch mechanism used in both language switching and task-switching, and perhaps best measured at long CTI (assuming Yehene & Merian, 2007 were correct that these pinpoint switch processes more specifically). To further investigate the absence of cross-domain correlations in switch costs, we combined bilingual and monolingual participants and recalculated cross-domain correlations collapsing together all 160 participants, as well as Spearman-Brown boosted split-half reliabilities comparing odd versus even blocks. The latter revealed significant correlations between odd and even blocks in switching costs at both long CTI (r = .29) and short CTI (r = .31, both ps < .05) in the color-shape task (see also Friedman & Miyake, 2004; Yehene & Meiran, 2007; Paap & Sawi, 2016), and in the language task (both rs = .37; ps < .01). Critically cross-domain correlations remained not significant even with all subjects combined (rs < .10).

However, our observation of robust correlations in mixing costs across-domains for both bilinguals *and* monolinguals, imply some significant limitations in the use of such correlations to test the hypothesis of shared control mechanisms.

Failures to Replicate the Bilingual Switching Advantage

The current analysis suggests that failed attempts to replicate the originally reported bilingual switching advantage (Prior & MacWhinney, 2010) should be attributed to relatively low reliability of switch cost measures, and the use of relatively short CTIs in the majority of these attempts. Additional, and possibly equally or even more important considerations include the use of non-overlapping button-response mappings, failure to consider practice effects, and failure to incorporate analysis of accuracy (which can be especially problematic in studies of individual differences; Hughes et al., 2014; Friedman & Miyake, 2004; Friedman et al., 2006; Miyake et al., 2000). Error rates, even if very low overall, may be deemed irrelevant without good reason (see Table 1 for summary of studies that did not report error analyses). In the current study, even at high levels of accuracy we found some significant effects in accuracy measures (see Figure 2). The bilingual switching advantage may also be restricted to, or more likely to appear in, bilinguals who switch languages frequently or active bilinguals (e.g., de Bruin et al., 2015; Hartanto & Yang, 2016; Prior & Gollan, 2011; Verreyt et al., 2015). These factors add to a long and rapidly growing list that might be critical to consider in future studies including variation in types of bilinguals studied, methodological differences in tasks, sample sizes, and approach to analysis (Paap, 2014; Paap & Sawi, 2014; for review see Paap, 2016), and even differing interpretations of relatively similar data patterns (compare de Bruin et al., 2015[8] versus Prior & Gollan, 2011).

---

[8] De Bruin et al., (2015) found significantly smaller switch costs for bilinguals than for monolinguals but argued against the notion of a bilingual advantage, because bilinguals tended to respond more slowly (albeit not significantly), and because when switch costs were calculated proportional to overall speed, the between-group difference in switch costs was no longer significant. In the current study, bilinguals did not respond significantly more slowly than monolinguals on the color-shape task. To match de Bruin et al's analysis, when we did control for baseline speed by calculating proportional switching cost (difference between switch and stay trials/stay trials * 100) our bilinguals exhibited a proportional switching cost that remained significantly smaller than that of monolinguals ($p = .02$) at long CTI in the first half of trials, consistent with the non-adjusted analyses. As in the current study, de Bruin et al., found that bilinguals tended to respond more slowly than monolinguals on stay trials, a pattern which led them to suggest that between-group differences might have reflected a difference in response strategy. Indeed, given the number of studies that have now reported this pattern, along with Prior & Gollan's (2011) report of a similar pattern in the language switching task for Spanish-English relative to Mandarin-English bilinguals, it might seem that bilingualism elicits a generally more cautious approach to mixed-task blocks. However, we offered a different interpretation in the present study based on separate examination of group comparisons on switch versus stay trials. Such comparisons might be further informative when trying to understand between group differences.

Another critical issue when measuring switching ability is reliability of the switching measures, and the extent to which they do or do not measure how switches, both linguistic and nonlinguistic, are planned and executed in daily life. Indeed, our finding of significant cross-domain correlations in mixing costs in monolinguals suggests substantial limitations on the extent to which the tasks used in the present study measure what bilinguals do to control activation of their two languages in daily life. Bilinguals' ability to juggle two languages involves many cognitive processes that are not captured by these type of switching tasks.

In our study, standardized clinical measures of switching ability also did not exhibit strong correlations with trial-to-trial task switching (and were instead better correlated with mixing costs; see Tables 5-6). Some CWIT measures were correlated with language mixing costs in bilinguals but not in monolinguals, perhaps implying some cross-domain transfer but not in a manner that leads bilinguals to consistently perform better on the CWIT (see Tao et al., 20105; but see Table 2) or mixing in general. Together these considerations suggest that better measures of both switching ability, and what ties those measures very specifically to what individuals do in daily life will be useful if they can be developed (perhaps along lines suggested by Green & Abutalebi, 2013; Hartanto & Yang, 2016; Yang, Hartanto, & Yang, 2016; Verreyt et al., 2015; Festman & Münte, 2012).

Why is the Group Difference in Switching but not Mixing Costs?

Our finding of robust correlations between mixing costs across domains (see also Prior & Gollan, 2013) implies overlap in the cognitive mechanisms underlying task-mixing and language-mixing and raises questions regarding why we did not observe a bilingual advantage in mixing costs. An important consideration here is that mixing costs appear to be more reliable than switching costs in both linguistic and non-linguistic switching paradigms. In the present study, mixing costs were also often significantly correlated with other measures, and others have reported higher test-retest reliability for mixing than switching costs in the color-shape task (Paap & Sawi, 2016). Indeed, perhaps it is not so surprising to find that that between-block differences appear to be more stable and reliable than within-block differences—mixing costs in the color-shape task are typically larger than switching costs (Rubin & Merian, 2005; but see Experiment 3 of Hernández et al., 2013). Additionally, as noted above, the finding that mixing costs were correlated across domains to the same extent in monolinguals as in bilinguals, implies that whatever overlap exists in control mechanisms used across tasks should not be attributed to bilingual language use. Thus, the absence of a bilingual advantage in mixing might simply reflect the fact that the tasks used herein do not adequately measure cross-domain transfer in mixing ability (assuming that such transfer does occur). This interpretation raises additional questions when considered in concert with our above interpretations of the switching advantage for bilinguals. Specifically, either cross-domain transfer in switching ability is stronger than cross-domain transfer in mixing, or the same tasks that measure cross-domain transfer in switching ability nevertheless do not measure cross-domain transfer in mixing ability – which seems a bit ad hoc.

In fact, some have suggested that mixing costs is where one would most expect bilingual advantages to arise—given that actual switches are relatively rare events, but the possibility to need to mix languages is always present when bilinguals are speaking (Prior & Gollan, 2013; Wiseheart et al., 2014; Bialystok et al., 2009 for review). Following this logic, Barac and Bialystok (2012) used the color-shape switching task with three groups of six-year old bilingual children (Chinese-English, French-English, and Spanish-English) and found smaller mixing, but not switching, costs for all bilingual groups relative to monolingual children. Possibly related, using language-switching to train color-shape switches and vice versa, Prior and Gollan (2013) found cross-domain transfer in mixing, but not switching, costs. On a related note, Soveri and colleagues (2011) found that a higher frequency of bilingual every-day language switching was related to a smaller *mixing* (but not switching) cost in errors on a number-letter switching task. Soveri and colleagues suggested that mixing cost reflects top-down management of two competing tasks, and thus closely resembles an everyday conversation in which a bilingual must decide which language to use by recruiting sustained control and general monitoring processes (although note that this finding was not replicated in a similar follow-up study from the same group; Jylkkä et al., 2017).

With the exception of the study by Barac and Bialystok (2012) and Gold and colleagues (2013), most of the color-shape switching studies that modeled Rubin & Meiran (2005) have not found a bilingual advantage in mixing (see Table 1). However, Wiseheart and colleagues (2014) hypothesized that the bilingual advantage should be in mixing costs in tasks that elicit interference between stimulus-response mappings. Thus, they used a paradigm in which stimulus-response mapping was changed on every trial (i.e., 100% response incompatible trials; see also Cepeda, Kramer, & Gonzalez de Sather, 2001). Specifically, in their study, the responses ‘blue’ and ‘horse’ were always indicated with the left index finger, and responses ‘red’ and ‘cow’ always with the right index finger. Additionally, only blue cows and red horses were presented as stimuli (no red cows and no blue horses); therefore, correct responding in this task required active updating of stimulus response associations on each and every switch trial (whereas in most switching studies with overlapping button response mappings a correct response could be produced 50% of the time for the wrong reason, i.e., even if a participant failed to switch tasks).

Most previous studies comparing bilinguals and monolinguals on task switching used non-overlapping responses with 50% of trials being response incompatible (see Table 1). Only a few studies used overlapping responses and also used 50% response incompatible trials (i.e., for half the trials the stimulus is associated with a different response; Garbin et al., 2010; Gold et al., 2013; the current study). Wiseheart and colleagues speculated that Gold and colleagues' finding of a mixing advantage for older adults may be explained by the fact that older adults are most sensitive to remapping and thus may show an advantage even with a 50% response incompatibility rate, whereas for a bilingual advantage in younger adults, a higher rate of response incompatible trials may be required (as found in their study). In future studies, it may be useful to vary the proportion of incompatible trials to see if this modulates how bilingualism influences performance in switching paradigms.

Greater consideration of possible points of overlap between linguistic and non-linguistic task switching where between domain transfer in control mechanisms might be most likely to occur might also be helpful. That is, what aspect of bilingual language use might benefit performance on a color-shape switching task given that these are obviously much more different than they are similar? It has been suggested that mixing costs reflect global conflict monitoring (Koch, Prinz, & Allport, 2005; Los, 1996, 1999), whereas switching costs reflect task reconfiguration, and the ability to inhibit the previous task in order to activate the currently relevant task (Meiran et al., 2000; Meuter, 2005; Monsell, 2003; Prior & MacWhinney, 2010; Prior & Gollan, 2013; Rubin & Meiran, 2005).

Initiation of some types of language switches might not involve domain-general mechanisms whatsoever (e.g., Kleinman & Gollan, 2016; Gollan & Goldrick, 2016; submitted). In the linguistic domain, languages are also naturally competing responses, whereas features such as color or shape are usually not in competition. Potential for overlap may come from the need in both domains to disengage from a prior task in order to engage with the new task, or from anticipating the need to switch when different cues are present (though in naturalistic settings the cues themselves would of course be very different, e.g., a person's face, a red or green traffic light). The present results suggest that preparation to switch might be an important factor mediating cross-domain transfer in control mechanisms, and that besides switching itself, other processes might ultimately be implicated (including preparation/planning, monitoring, in addition to switching itself). But the absence of bilingual advantage in mixing, and the absence of consistent advantages in switching, remain a bit of a mystery at this point in time.

Concluding remarks

Investigation of bilingual advantages in task switching was triggered by the assumed existence of a general switching ability, and the possibility of transfer between linguistic and nonlinguistic processing domains, as well as a perhaps implicit assumption that frequent switchers should exhibit characteristics potentially associated with "elite" switching ability. However, the pursuit of such ideas can easily and quickly be misleading, carrying with them untested assumptions that may or may not be correct. Here we found evidence for switching ability as a skill that improves with practice, and ample preparation time may allow this general ability (as well as its role in both task and language switching) to be revealed in trial-to-trial switching as measured in the lab with experimental control. Paradoxically, practice reduced switching costs the most with*out* ample preparation time (i.e., at short CTI), thus further work remains to be done to explore how preparation time, practice, and cross-domain transfer might interact (specifically, it is not clear that task-specific practice and cross-domain transfer should be explained in the same way; see also Prior & Gollan, 2013). Expectations that highly practiced switchers should be advantaged especially on switch trials can also be misleading, and must be considered relative to an appropriate baseline, with joint consideration of error rates, and other factors that may introduce between group differences. A pattern that by direct comparison seems to implicate one condition in explaining between groups differences, might be viewed differently when performance is considered across different trial types within each group. As a field, development of measures that are more clearly tied to naturally occurring experience with greater specification of what exactly should improve with practice will likely be fruitful both for understanding bilingual language control and more broadly for understanding task switching and multi-tasking.

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*, 242–275.

Abutalebi, D. J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, *23*, 557–582.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., Cappa, S., & Costa, A. (2011). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, *22*, 2076–2086.

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: role of language, cultural background, and education. *Child Development*, *83*, 413–422.

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics, 35*, 857-894.

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, *65*, 229–235.

Bialystok, E., Craik, F.I.M., Green, D.W., & Gollan, T.H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10,* 89–129.

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2015). Language control in bilinguals: Monitoring and response selection. *Cerebral Cortex*, 1–14. doi:10.1093/ cercor/bhv052

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*, 65–78.

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*, 278–288.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*, 282–298.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. (2001). Changes in executive control across the life span: examination of task-switching performance. *Developmental Psychology, 37*, 715–730.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*, 135–149.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50,* 491–511.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*, 336–345.

Deák, G. O., & Wiseheart, M. (2015). Cognitive flexibility in young children: General or task-specific capacity? *Journal of Experimental Child Psychology*, *138*, 31–53.

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*, 1752–1765.

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.

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, *22*, 1630–1645.

Delis, D.C., Kaplan, E., & Kramer, J.M. (2001). Delis-Kaplan Executive Function System (D-KEFS) examiner's manual. San Antonio, TX: The Psychology Corporation.

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

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Festman, J., & Münte, T. F. (2012). Cognitive control in Russian-german bilinguals. *Frontiers in Psychology*, *3*, 1-7.

Fink, A., & Goldrick, M. (2015). Pervasive benefits of preparation in language switching. *Psychonomic Bulletin & Review*, *22*, 808–814.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology. General*, *133*, 101–135.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172–179.

Gade M., Koch I. (2007). The influence of overlapping response sets on task inhibition. *Memory & Cognition. 35,* 603–609.

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., Hernandez, M., Costa, A., & Avila, C. (2010). Bridging language and attention: brain basis of the impact of bilingualism on cognitive control. *NeuroImage*, *53*, 1272–1278.

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. *Journal of Neuroscience*, *33*, 387–396.

Gollan, T.H., & Goldrick, M (submitted). A switch is not a switch: Syntactically-driven bilingual language control.

Gollan, T. H., Kleinman, D., & Wierenga, C. E. (2014). What's easier: doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General*, *143*, 2167–2195.

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science*, *22*, 1155–1164.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). 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*, 594–615.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81.

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

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature Reviews. Neuroscience*, *11*, 651–659.

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: developmental trajectories and mediation. *Developmental Science*, *18*, 686–702.

Hartanto, A., & Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: A diffusion model analysis of the effects of compound bilingualism 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? *Cambridge Handbook of Bilingual Processing*. Cambridge: CUP.

Houtzager, N., Lowie, W., Sprenger, S., & de Bot, K. D. (2015). A bilingual advantage in task switching? Age-related differences between German monolinguals and Dutch-Frisian bilinguals. *Bilingualism: Language and Cognition*, 1–11.

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: their reliability and increased validity. *Behavior Research Methods*, *46*, 702–721.

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

Kiesel, A., Steinhauser, M., Wendt, M., Falkensten, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching: A review. *Psychological Bulletin, 136*, 849-874.

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. *The Quarterly Journal of Experimental Psychology, 63*, 595-619.

Kleinman, D., & Gollan, T. H. (2016). Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological Science, 27,* 700-714.

Koch, I., Prinz, W., & Allport, A. (2005). Involuntary retrieval in alphabet-arithmetic tasks: Task-mixing and task-switching costs. *Psychological Research*, *69*, 252–261.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology, 25*, 497-514.

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

Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, *94*, 145–188.

Los, S. A. (1999). Identifying stimuli of different perceptual categories in mixed blocks of trials: Evidence for cost in switching between computational processes. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 3–23.

Ma, H., Hu, J., Xi, J., Shen, W., Ge, J., Geng, F., Wu, Y., Guo, J., & Yao, D. (2014). Bilingual cognitive control in language switching: An fMRI study of English-Chinese late bilinguals. *PLoS ONE*, *9*, 1-8.

Maxwell, S. E. and Delaney, H. D. (2004). Designing experiments and analyzing data: a model comparison perspective. Lawrence Erlbaum Associates, Mahwah, N.J., 2nd edition.

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and Aging*, *16*, 96–109.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29,* 362–372.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*, 211–253.

Meuter, R. (2005). Language selection in bilinguals: Mechanisms and processes. In J.F. Kroll & De Groot, A.M.B Eds. Handbook of Bilingualism: Psycholinguistic approaches, New York, NY, Oxford University Press, pp 349-370.

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

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychology*, *21*, 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.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140.

Mor, B., Yitzhaki-Amsalem, S., & Prior, A. (2015). The joint effect of bilingualism and ADHD on executive functions. *Journal of Attention Disorders*, *19*, 527–541.

Myers-Scotton, C. M. (2006). Natural codeswitching knocks on the laboratory door. *Bilingualism: Language and Cognition*, 9, 203–212.

Paap, K. R. (2014). The role of componential analysis, categorical hypothesising, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology*, *26*, 242–255.

Paap, K.R. (2016). The neuroanatomy of bilingualism: will winds of change lift the fog? *Language, Cognition and Neuroscience, 31*, 331-334.

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. (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., Myuz, H.A., Anders, R.T., Bockelman, M.F., Mikulinsky, R., & Sawi, O.M. (2016). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology, 29,* 89-112.

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Language Sciences*, *5*, 962.

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.

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 1-8.

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-define response sets. *European Journal of Cognitive Psychology, 19*, 395–416.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society*, *17*, 682–691.

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*, 622–645.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*, 253–262.

Reitan, R.M. (1992). Trail Making Test Manual for administration and scoring. South Tucson, AZ: Reitan Neuropsychology Laboratory.

Rodríguez-Pujadas, A., Sanjuán, A., Ventura-Campos, N., Román, P., Martin, C., Barceló, F., Costa, A., & Ávila, C. (2013). Bilinguals use language-control brain areas more than monolinguals to perform non-linguistic switching tasks. *PLoS ONE,* 8, 1-8.

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

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within-group analysis approach. *Frontiers in Psychology*, *2,* 1-8.

StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643.

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.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*, 77–83.

Turrell, G., Lynch, J. W., Kaplan, G. A., Everson, S. A., Helkala, E.-L., Kauhanen, J., & Salonen, J. T. (2002). Socioeconomic position across the lifecourse and cognitive function in late middle age. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 57,* S43–51.

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*, 3–24.

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition, 110,* 84–99.

Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, *19*, 181–190.

Wagenmakers, E.J. (2015). A quartet of interactions. *Cortex, 73*, 334–335.

Weissberger, G. H., Gollan, T. H., Bondi, M. W., Clark, L. R., & Wierenga, C. E. (2015). Language and task switching in the bilingual brain: Bilinguals are staying, not switching, experts. *Neuropsychologia*, *66*, 193–203.

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.

Williams, V. S. L., L. V. Jones, and J. W. Tukey. (1999). "Controlling Error in Multiple Comparisons, with Examples from State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioral Sciences*, *24*, 42–69.

Winer BJ, Brown DR, Michels KM. Statistical principles in experimental design. New York: McGraw-Hill; 1971.

Wiseheart, M., Viswanathan, M., & Bialystok, E. (2014). Flexibility in task switching by monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *19*, 141–146.

Yang, H., Hartanto, A., & Yang, S. (2016). The complex nature of bilinguals' language usage modulates task-switching outcomes. *Frontiers in Psychology*, *7*, 1-8.

Yehene, E., & Meiran, N. (2007). Is there a general task switching ability? *Acta Psychologica*, *126*, 169–195.

Authors' Note

We thank Mayra Murillo Beltran, Beatriz Bobadilla, Rosa Elizalde, and Cecilia Salcedo for their help with data collection and Daniel Kleinman and Victor Ferreira for helpful discussion. The research reported here was funded by grants from the National Institute on Deafness and Other Communication Disorders (011492), National Institute of Child Health and Human Development (050287, 051030, 079426), and by Graduate Research Fellowship Program (GRFP) from the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NIH or NSF. The data reported here were previously presented as a poster at Psychonomic Society's 57th Annual Meeting in Boston, MA in November of 2016. There were no conflicts of interest in completing this research.

Figure Captions

Figure 1. Reaction times (in ms) for single, stay, and switch trials in the color-shape task shown for the first half (1a) and second half (1b) of experimental trials for both language groups. The asterisk denotes a significant difference in switching costs between the bilingual and monolingual groups at $p < .05$.

Figure 2. Error rates (in % incorrect) for single, stay, and switch trials in the color-shape task for the first half (2a) and second half (2b) of the experimental trials for both language groups. Asterisks denote significant differences in error rates between the bilingual and monolingual groups at $p < .05$.

Figure 3. Mean difference scores (i.e., switching and mixing costs in ms) broken down by first and second half of experimental trials for each CTI for both language groups.

Figure 4. Reaction times (in ms) for single, stay, and switch trials in the language task shown for the first half (5a) and second half (5b) of experimental trials for the bilingual group. Reaction times from the color-shape task are shown for comparison (same means as shown in Figure 1).

Figure 5. Across-domain and within-domain correlations between color-shape and language tasks. Bilinguals' data shown in black dots and black lines, and monolinguals' data in gray triangles and gray lines.

Figure 6. The correlation between socio-economic status (SES, as measured by average parental education) and switching and mixing costs in the color-shape task by language group. Bilinguals' data shown in black dots and black lines, and monolinguals' data in gray triangles and gray lines.