Preparing to switch languages versus preparing to switch tasks: which is more effective?

Brontë Graham and Aureliu Lavric

Department of Psychology, University of Exeter, UK

(accepted for publication in Journal of Experimental Psychology: General)

DOI: https://doi.org/10.1037/xge0001027

Corresponding author:

Professor Aureliu Lavric
Psychology
College of Life and Environmental Sciences
Washington Singer Laboratories
University of Exeter
Exeter EX4 4QG
UK
Tel: +44 (0)1392 724642
E-mail: a.lavric@exeter.ac.uk

Author Note

We are grateful for comments on and discussions of earlier versions of this manuscript to Stephen Monsell, Heike Elchlepp, Tobias Stevens, Ian McLaren, Nicolas Dumay, Andrea Philipp and two anonymous reviewers. The results of this study were presented as posters at the 2019 Annual Meeting of the Psychonomic Society in Montreal, and at the 21st Conference of the European Society for Cognitive Psychology (ESCoP) in 2019 in Tenerife. The research data from this study will be made available on Open Science Framework, https://osf.io/9haqg/.

ABSTRACT

A substantial literature relates task-set control and language selection in bilinguals – with "switching" paradigms serving as a methodological "bridge". We asked a basic question: is preparation for a switch equally effective in the two domains? Bilinguals switched between naming pictures in one language and another, or between the tasks of naming and categorizing pictures. The critical trials used for comparing the two kinds of switching were identical in all respects – task (naming), stimuli, responses – except one: whether the shape cue presented before the picture specified the language or the task. The effect of preparation on the "switch cost" was examined by varying the cue-stimulus interval (CSI=50/800/1175 ms). Preparation for a task switch was more effective: increasing the CSI from 50 to 800 ms reduced the RT task switch cost by ~63% to its minimum, but the language switch cost only by ~24%, the latter continuing to reduce with further opportunity for preparation (CSI=1175 ms). The switch costs in the two domains correlated moderately (r = .36). We propose that preparation for a language switch is less effective, because (a) it must pre-emptively counteract greater interference during a language switch than during a task switch, and/or (b) lexical access is less amenable to "top-down" control than (components of) task-set. We also investigated the associations between stimuli and the language (or task) where they were last encountered. Associative history influenced performance – but similarly for switches and repetitions – indicating that stimulus-induced associative retrieval of language (or task-set) did not contribute to switch costs.

A little over two decades ago researchers interested in intentional control of attention and performance began using a paradigm where participants were first familiarised with two (or more) simple cognitive tasks and then asked to switch between these tasks on demand (Rogers & Monsell, 1995; Meiran, 1996; Allport, Styles, & Hsieh, 1994; see also Jersild, 1927; Shaffer, 1965; Spector & Biederman, 1976; Sudevan & Taylor, 1987). Thereafter, this "task switching" paradigm gained considerable popularity (for reviews see Kiesel et al., 2010; Monsell, 2003; 2015; 2017; Vandierendonck, Liefooghe, & Verbruggen, 2010), yielding important insights and several intriguing empirical phenomena, including, but not limited to, the robust performance "switch cost" (Rogers & Monsell, 1995). Conveniently, the paradigm could be easily adapted for investigating different cognitive processes by requiring a switch in some, but not other, components of the task-set. For example, if one is interested in the intentional control of attention to perceptual input, one can design "attention switching" paradigms – where the relevant perceptual attribute is the only aspect of the task that changes. This can be the relevant visual dimension (e.g., colour vs. shape, e.g., Meiran & Marciano, 2002), location (Longman et al., 2014, 2016, 2017), perceptual modality (vision vs. audition, Lukas, Philipp, & Koch, 2010), or voice in a multitalker compound (e.g., Koch, Lawo, Fels, & Vorlander, 2011; Monsell, Lavric, Strivens & Paul, 2019), whist other parameters of the task-set (e.g., the required categorization, S-R rules) can be kept constant. Similarly, if one is interested in how bi(multi)linguals intentionally select the language for production, one can isolate the output language as the crucial task-set component, and require bilingual participants to switch languages keeping other task requirements (e.g., to name the displayed picture) constant from one trial to the next.

This "language switching" variant of the paradigm emerged shortly after the above-mentioned task switching variant (e.g., Meuter & Allport, 1999), and has since become one of the most popular paradigms in the bilingualism literature (see Declerk & Philipp, 2015, for a review). Its conceptual and procedural similarity to the task switching variant made it the paradigm of choice for researchers investigating whether control processes at play in task switching and language switching are predominantly common/shared ("domain-general") or predominantly "domain-specific". The interest in this issue has arisen at least in part from the attractive (e.g., Titone & Baum, 2014), yet controversial (Bialystok & Craik, 2015; Hilchey, Saint-Aubin, & Klein, 2015; Kroll & Bialystok, 2013; Paap & Greenberg, 2013), notion that extensive day-to-day language selection results in superior control of other cognitive processes in bilinguals relative to monolinguals, and even makes bilinguals more resilient to the effects of neurodegeneration (e.g., Bialystok, Craik, & Freedman, 2007).

To this end, a substantial body of research has examined the degree of overlap between empirical phenomena documented in task switching and language switching. A succinct tentative summary of the relevant phenomena must start with the (perhaps unsurprising) ubiquity of the "switch cost" in both domains – performance is invariably poorer when one changes task or language compared to staying in the same task or language[1]. A pertinent question is whether the switch costs in the two domains correlate over individuals. Most studies that examined this correlation found it to be weak and non-significant (Branzi, Calabria, Boscarino, & Costa, 2016; Calabria, Hernández, Branzi, & Costa, 2012; Klecha, 2013; Prior & Gollan, 2013; Stasenko, Matt, & Gollan, 2017), or significant yet still modest (Timmer, Calabria, Branzi, Baus, & Costa, 2018), thus casting doubt on the commonality of the sources of the task switch costs and language switch costs. However, recently Declerck, Grainger, Koch, and Philipp (2017) have argued that the apparent lack of correlation is due to the large methodological discrepancies between the linguistic and non-linguistic tasks being compared in these experiments – differences in stimulus type (e.g., alphanumeric characters vs. images), number of stimuli, tasks (e.g., naming vs. categorization), number of response alternatives and response modality (keypress vs. speech). Declerck et al. (2017) have re-visited the correlation whilst carefully matching the two variants of switching paradigm. They started by matching the type and number of stimuli, and the number of responses and their modality (but not the task). In two subsequent experiments, they matched all of the above parameters, including the task. All three experiments yielded considerably higher correlations than those reported hitherto: $r = .44$ when the task was not matched; $r = .57$ and $r = .64$, when it was (all statistically significant).

Another parallel drawn between the two domains concerns the "paradoxical asymmetry" of switch costs. If one is asked to switch between a less (or less recently) practiced task and a more (or more recently) practiced task, the switch cost is typically larger for the latter ("stronger") task than the former ("weaker") task (e.g., Allport, Styles, & Hsieh, 1994; Yeung & Monsell, 2003). Similarly, larger switch costs for the "stronger" language (in which the bilingual has had more practice) compared to the "weaker" language have also been reported (e.g., Meuter & Allport, 1999). However, in language switching this

[1] Two notable exceptions are: the recently reported cost-free voluntary language switching in bilinguals (Kleinman & Gollan, 2016, see below for further description of this study), and the negligibly small (and non-significant) task switch cost reported by Verbruggen, Liefooghe, Vandierendonck, and Demanet (2007) when task cues were presented briefly and removed from the display before the stimulus was presented (though our own subsequent attempts to follow this "recipe" of brief cue presentation did not result in a near-elimination of the task switch cost, e.g., Longman et al., 2014).

"paradoxical asymmetry" of switch costs has been much less consistent than in task switching. For example, Calabria et al. (2012) found that bilinguals who showed a robust and persistent "paradoxical" asymmetry of task switch costs did not show comparable asymmetry of language switch costs, only a weak trend towards "paradoxical" asymmetry in the first half of the testing session which was reversed in the second half of the session.

Among the phenomena compared in the two domains there is also the "mixing cost" – poorer performance on task (or language) repetition trials in blocks containing switch trials compared to single-task (or single-language) blocks (e.g., Los, 1996; Stasenko et al., 2017). A recent study with a relatively large N (Stasenko et al., 2017) found a moderate ($r \sim .4$), and statistically significant, correlation between the task mixing cost and the language mixing cost despite some large differences between the paradigms (in stimuli, tasks, etc.). A further phenomenon worth mentioning is the n-2 repetition cost, also referred to as "backward inhibition" (Mayr & Keele, 2000) – in a sequence of at least 3 tasks containing no task repetitions, performance is worse when one switches back to the task performed 2 trials ago relative to switching to another task (e.g., for tasks A, B and C, performance is worse on the third trial in the sequence ABA than in the sequence CBA). The n-2 repetition cost has also been consistently observed in language switching (e.g., Philipp, Gade, & Koch, 2007; Philipp & Koch, 2009). Yet, studies which examined the n-2 repetition costs in language switching and task switching (Branzi et al., 2016; Timmer et al., 2018) found them to be uncorrelated. Recent studies have also compared the effects of voluntary language (or task) choice on the language switch cost and the task switch cost (e.g., Gollan, Kleinman, & Wierenga, 2014; Kleinman & Gollan, 2016). One intriguing finding in the voluntary switching literature is that language switching can be cost-free when bilinguals are allowed to choose the language in which they prefer to name each picture, presumably because this minimizes the need for top-down control (Kleinman & Gollan, 2016), whereas switching tasks comes at a cost even when they are voluntary (we are not aware of any evidence of cost-free voluntary task switching). Another phenomenon examined across the two domains is the effect of aging on the switch costs and mixing costs (Gollan & Ferreira, 2009; Weisberger, Wierenga, Bondi, & Gollan, 2012). So far, the evidence suggests mostly differential effects of aging on task switching vs. language switching and possibly greater resilience of the latter – for example, old age bilinguals who cannot switch between tasks when required seem nevertheless able to switch between languages on demand (Weisberger et al., 2012).

Most of the research reviewed above has been motivated by the "domain-general vs. domain-specific control" framework, which has been pivotal in the theoretical and empirical characterization of intentional control of language selection and task-set selection. Our current emphasis is somewhat different – it is primarily on *the effectiveness* of intentional control in these two domains. Giving participants time to prepare for the upcoming task (e.g., by presenting the task cue in advance of the imperative stimulus, Meiran, 1996) tends to substantially reduce the task switch cost (e.g., Rogers & Monsell, 1995; Meiran, 1996; Monsell & Mizon, 2006). In the task switching literature this reduction in switch cost with preparation (RISC effect) is widely considered the clearest index of intentional task-set control, because it cannot be attributed to processes elicited exogenously by the stimulus (e.g., Meiran, 1996; Monsell, 2003). Whilst task switching studies have reported consistent and robust RISC effects (see Monsell, 2015; 2017, for reviews), including studies that involved linguistic tasks, such as lexical or semantic decision about words (e.g., Elchlepp, Lavric, & Monsell, 2015), in language switching the evidence for the RISC effect is mixed. Some studies have reported robust RISC effects (Costa & Santesteban, 2004; Fink & Goldrick, 2015; Mosca & Clahsen, 2016), but other studies found no detectable effect of preparation on the switch cost (Stasenko et al., 2017), or only a modest RISC effect limited to some conditions (Declerck, Ivanova, Grainger, & Duñabeitia, 2020; Lavric, Clapp, East, Elchlepp, & Monsell, 2019), or even an increase in switch cost with preparation (Philipp et al., 2007). Might the volatility of the reduction in the language switch cost be due to methodological limitations? Task switching research has documented two potentially serious confounding factors that tend to inflate the RISC effect:

- First, in task cuing (the most widely used variety of task switching paradigm first introduced by Meiran, 1996, where a cue specifies the task on each trial), using a single cue per task is problematic, because on task repetition trials the cue is always the same as on the previous trial, whereas on task switch trials the cue always changes. This results in an extra benefit for cue encoding on task repeat relative to task switch trials. This effect (to which we henceforth refer to as the "cue change/repetition" effect) has been shown to confound (inflate) both the switch cost and its reduction with preparation (e.g., Logan & Bundesen, 2003; Mayr & Kliegl, 2003; Monsell & Mizon, 2006). Recently cue change/repetition was also shown to inflate the language switch cost (Heikoop, Declerck, Los, & Koch, 2016, though this study did not examine the effect of cue change/repetition on the RISC). None of the above-mentioned language switching studies that reported significant RISC effects (Costa & Santesteban, 2004; Declerck et al., 2020; Fink & Goldrick, 2015; Mosca & Clahsen, 2016) have addressed the cue change/repetition confound. A common solution in task switching is to never repeat the cue from the last trial, not even when the task is repeated (evidently, this requires a minimum of 2 cues per task). The only language switching study (to our knowledge) that controlled for the effect of cue change/repetition (Lavric et al., 2019) found a RISC effect only for one of the two types of cue – and there it was modest and statistically significant only for median (but not mean) RTs. Thus, it seems that the RISC effects reported in earlier studies may have been inflated by repeating the cue on language repetition trials.

- Second, the estimation of the RISC effect is based on manipulating the time available for preparation. Yet, one must ensure that increasing the preparation interval does not also increase the time elapsed from the previous response – the response-stimulus interval. Increasing the latter has been shown to reduce the switch cost independently of preparation (Meiran, 1996), possibly because it provides more opportunity for passive dissipation of the "task-set inertia" from the previous trial. Of the four language switching studies that have ensured that preparation is not confounded by the interval from the previous response – only one reported a robust RISC effect (Mosca & Clahsen, 2016), whereas in the other three the RISC effect was either modest (and confined to one language, Declerck et al., 2020, or one type of cue, Lavric et al., 2019), or altogether absent (Philipp et al., 2007).

Thus, careful examination of the RISC effects in the language switching literature suggests that when crucial confounds are addressed, preparation may be less effective in reducing the language switch cost than previously thought or assumed. Numerous task switching studies, which have addressed the above confounds in the same way as Lavric et al.'s (2019) language switching study, have reported large and statistically robust RISC effects (e.g., Monsell & Mizon, 2006; Lavric, Mizon, & Monsell, 2008; Longman, Lavric, Munteanu & Monsell, 2014; Van't Wout, Lavric, & Monsell, 2013; 2015). Since the RISC effect is seen as a "litmus test" of intentional (top down) control, the implication is that intentional control – specifically, preparatory control – may be less effective in language selection than in task-set selection. At first glance, this may seem counterintuitive – indeed, outside the laboratory bilinguals seem to switch languages effortlessly. However, in doing so they likely exploit a variety of contextual and other cues that may help activate the relevant language exogenously – hence intentional ("top-down") control of language selection may not be always (or often) required. Moreover, there may be other plausible reasons for language selection being less amenable to effective preparatory control. The links between meaning and output phonology in each language are very strong, and there is evidence that meaning simultaneously activates the early stages of output for both languages (e.g., Chabal & Marian, 2015). This means that preparatory selection (activating the target language before the to-be-communicated meaning is determined) may not be very effective at preventing the meaning (once determined) activating phonological representations in the non-target language. Indeed, influential theories of bilingual control such as Greene's (1998) Inhibitory Control model postulate a key role for late (post-stimulus, *reactive*) control processes in resolving the conflict between the competing outputs. In contrast, task switching studies use novel tasks and typically rely on newly-acquired, arbitrary, associations between stimuli and responses. In these circumstances preparatory selection of the relevant set of S-R mappings may be rather effective in preventing (or strongly reducing) the activation of irrelevant S-R mappings when the stimulus becomes available to perception.

Intriguing as these considerations may be, one must first obtain firm evidence by directly comparing the effectiveness of preparatory control in the two domains. The current study aims to do so – it examines the RISC effect by manipulating the preparation interval in bilingual participants required to switch tasks or languages (in separate sessions). Crucially, we compare the two domains using the same task (picture naming in L1 and L2), and identical cues, stimuli, and responses. The only element differentiating the two variants of the paradigm is that in the task switching variant participants switched to picture naming from another linguistic task (picture categorization), whilst in the language switching variant participants switched to picture naming in one language from picture naming in another language. Furthermore, the design of the study was optimized based on the aforementioned (and other) key "lessons" from the task switching literature. First, we unconfounded the effect of task (or language) change from the effect of cue change, by ensuring that the language (or task) cue always changed from one trial to the next, even when the language (or task) was repeated. Second, we unconfounded the effect of preparation from any effects of the interval elapsed from the previous response by ensuring that changes in preparation (cue-stimulus) interval did not systematically influence the response-stimulus interval. Third, previous research has shown that when the proportion of task switches is relatively high (≥50%) participants anticipate a possible (likely) switch before the presentation of the task cue, which results in underestimation of switch costs and RISC effects (Monsell & Mizon, 2006; Mayr, Kuhns, & Rieter, 2013; Kikumoto, Hubbard, & Mayr, 2016). To discourage participants from adopting such strategies, and thus maximise sensitivity to the RISC effect, we used a relatively low proportion of switch trials (33%). Fourth, by using three preparation intervals (most language switching studies to date have used two), we aimed to provide a better characterization of the preparation function and potentially capture its asymptote (which requires sampling a minimum of three points). Finally, we ensured more than sufficient power for detecting a difference between the RISC effects in task switching vs. language switching, by testing a sample of 48 bilinguals, and ensuring 68 trials per cell per participant for the within-participants analysis of the relevant statistical interaction involving factors *switch/repeat*, *preparation interval* and *language switching/task switching variant*. A "rule-of-thumb" recommendation in a recent analysis of power in cognition experiments is that an adequately powered experiment requires a total of 1600 observations (participants x trials) per cell of the relevant analysis (Brysbaert & Stevens, 2018). In our design the total number of observations per cell (for the above interaction) is 3072 (48 participants x 68 trials), nearly double the number recommended by Brysbaert and Stevens. Thus, even if one accounts for some attrition (e.g., error trials in RT analyses), the study is more than adequately powered (see Method for further power analyses).

Although the primary objective of our study is to compare how effective preparation is in reducing the task switch cost vs. the language switch cost, our paradigm also enables us to examine two other phenomena that are of considerable interest. Both pertain to the issue of the commonality of the "sources" of the language switch costs and task switch costs. The first is the correlation between the switch costs in the two domains. If there is at least some overlap in the sources of the switch cost, one would expect at least a moderate correlation. As already mentioned, several previous studies found this correlation to be small and (mostly) non-significant, but recently Declerck et al. (2017) found that a closer match of the paradigms results in larger, and statistically significant, correlations. Unfortunately, Declerck and colleagues have not unconfounded their switch cost measure from the effect of cue change/repetition (see above) – which could mean that cue change/repetition constituted a nontrivial portion of their task switch cost and their language switch cost. Hence, the correlations they have reported could reflect the (interesting) overlap between control processes in the two kinds of switching, or the (uninteresting) overlap between the facilitation of cue encoding in the two domains, or, more likely, some combination of these. The current study can help resolve this ambiguity, because, like Declerck et al.'s (2017) study, we matched rigorously the language switching vs. task switching paradigms – yet, we also controlled for cue change/repetition.

Second, if one is to examine the overlap in the sources of the switch costs in the two domains, one should investigate whether theoretical accounts of the sources of switch costs developed in one domain apply in the other domain. Our study may be able to test in language switching an influential account of the source of the task switch cost. It posits that over the course of the experiment each stimulus forms associative bindings with each task-set in whose context the stimulus is encountered. Thus, a stimulus may automatically retrieve ("re-activate") task-sets via such associative bindings, resulting in facilitation and/or interference, depending on the strength of the association between the stimulus and each task-set (Waszak, Hommel, & Allport, 2003; 2004; 2005). Importantly, Waszak and colleagues proposed that stimulus-elicited associative retrieval of the irrelevant task-set is more detrimental on a switch trial, where the activation of the relevant task-set (relative to the activation of the irrelevant one) is weaker. This "associative history" account of the task switch cost has received support from studies that compared performance for stimuli previously encountered only in the context of the currently relevant task vs. stimuli previously encountered in the context of both the currently relevant and in the context of the competing (currently irrelevant) task. Performance was found to be worse for the latter than the former – and, crucially, this "associative history" effect was greater when the task switched than when it was repeated, resulting in a larger switch cost for stimuli previously encountered in the context of the irrelevant task (e.g., Waszak et al., 2003). Here we employ a similar kind of analysis for language switching data (and for task switching data), by examining the history of the most recent encounter with each stimulus. We expect associative history to modulate the overall performance. More importantly, if the stimulus-elicited associative activation of the non-target language is a major contributor to the language switch cost, we expect the switch cost to be smaller on trials where the stimulus was most recently named in the same language as that required on the current trial, than on trials where the stimulus was most recently named in the other (currently irrelevant) language.

METHOD

Participants

Forty-eight bilinguals (12 male, 36 female; mean age=20.73; *SD*=2.16) whose first language was French (19), Spanish (17) or German (12) were recruited on the University of Exeter campus via opportunity sampling through social media publications and social networks, and via word of mouth. All participants provided informed written consent to participate in the study whose procedure adhered to the guidelines of (and was approved by) the local Ethics Committee (Psychology, University of Exeter); participation was remunerated with £20.

Our approach to determining the adequacy of our sample size was two-fold. First, as already explained in the Introduction, we have examined our sample size in the light of recent recommendations of Brysbaert and Stevens (2018) based on their analyses of a cognition mega-study, and found our sample size to be more than adequate. Second, we conducted a power analysis of our most recent 10 task switching experiments (9 of which have been published)[2]. Neither we (nor, to our knowledge, others) have directly compared the RISC effects in task switching and language switching, so we could not estimate based on the empirical data the sample size required for detecting the RISC x paradigm interaction. However, we had a good indication for the sample size required to detect a RISC effect in task switching experiments that used a very similar task cuing procedure to that employed in the current study. We used the G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) "a priori" procedure to estimate the sample size (N) given the expected effect size (based on our previous task switching studies), at the required α and power thresholds. This analysis yielded for $\alpha \leq 0.05$ a mean estimate of N = 9.7 (median 10, range 5-14) required to achieve power ≥ 0.8; and N = 11.9 (median 12, range 6-12) to achieve power ≥ 0.9. Thus, our N = 48 is 4-5 times larger, and hence more than adequate for detecting effects similar to (and likely substantially smaller than) those we found in our 10 previous experiments.

A questionnaire designed in-house was used to obtain the following details about the participants' linguistic background. Participants had lived in an English-speaking country for an average of 2.48 years (*SD*=2.01), and in the country of their native language for 17.07 years (*SD*=4.75), and had started acquiring English an average of 13.88 years (*SD*=3.99) before the study was conducted. Participants were asked to classify their level of proficiency in English using the Common European Framework of Reference for languages (CEFRL, Council of Europe, 2001), which provides qualitative descriptors for 6 levels of proficiency separately for listening, reading, spoken interaction, spoken production, and writing. To obtain a numerical summary of the self-assessed proficiency, we converted the 6 levels of proficiency into a 1-6 scale, where 6 indicated the highest proficiency. The average scores were: 5.58 (*SD*=.60) for listening, 5.54 (*SD*=.61) for reading, 5.35 (*SD*=.66) for spoken interaction, 5.33 (*SD*=.72) for spoken production, and 5.42 (*SD*=.64) for writing. The overall proficiency (averaging over the different kinds of activity) was 5.45 (*SD*=.56). Thus, self-

[2] A detailed description of these power analyses is available on Open Science Framework along with the data and materials from the study (see Author Note).

assessed proficiency, as well as the demographic and L2 acquisition data above suggest that our participants were highly proficient in English.

Tasks and materials

For each participant there were three testing sessions: one language switching session and two task switching sessions (one in L1 and one in L2). The order in which participants completed the sessions was counterbalanced so that half of the participants completed language switching followed by task switching and half did the opposite. Within each of these groups, the order of the two task switching sessions was counterbalanced: half started with the session in L1 session and half with the session in L2 (see Table 1). As illustrated in Figure 1, in the language switching session participants were required on every trial to name in either L1 or in L2 a greyscale drawing of an object or a person (mean dimensions: 39 x 46 mm, SDs: 13.3 mm and 7.4 mm) presented on a flat-screen monitor positioned at approximately 60 cm from their eyes. In the task switching sessions participants were required to either name the drawing or categorize it using a spoken response. The same 16 drawings were used in the language switching and task switching sessions. The drawings were selected so that they could be naturally classified into four semantic categories (4 drawings per category), for which participants were required to make the following responses: “clothes”, “job”, “food”, and “body” (the latter referring to the “body parts” category). The majority of the stimuli were the same across the 3 groups of bilinguals (French, Spanish and German), however there were some differences (see Table 2).

Table 1

The combined counterbalancing of: (1) the order of task switching vs. language switching, (2) the order of the two task switching sessions, and (3) the allocations of the two sets of cues (see Fig. 1, upper panel) to task switching vs. language switching.

Participant	Session 1	Session 2	Session 3
	Task switching (L1)	Task switching (L2)	Language switching
1	Cues for naming: heart, square Cues for categorizing: cross, hexagon	Cues for naming: heart, square Cues for categorizing: cross, hexagon	Cues for naming in L1: circle, diamond Cues for naming in L2: triangle, star

2	Task switching (L1)	Task switching (L2)	Language switching
	Cues for naming:	Cues for naming:	Cues for naming in L1:
	circle, diamond	circle, diamond	heart, square
	Cues for categorizing:	Cues for categorizing:	Cues for naming in L2:
	triangle, star	triangle, star	cross, hexagon
3	Task switching (L2)	Task switching (L1)	Language switching
	Cues for naming:	Cues for naming:	Cues for naming in L1:
	heart, square	heart, square	circle, diamond
	Cues for categorizing:	Cues for categorizing:	Cues for naming in L2:
	cross, hexagon	cross, hexagon	triangle, star
4	Task switching (L2)	Task switching (L1)	Language switching
	Cues for naming:	Cues for naming:	Cues for naming in L1:
	circle, diamond	circle, diamond	heart, square
	Cues for categorizing:	Cues for categorizing:	Cues for naming in L2:
	triangle, star	triangle, star	cross, hexagon
5	Language switching	Task switching (L2)	Task switching (L1)
	Cues for naming in L1:	Cues for naming:	Cues for naming:
	heart, square	circle, diamond	circle, diamond
	Cues for naming in L2:	Cues for categorizing:	Cues for categorizing:
	cross, hexagon	triangle, star	triangle, star
6	Language switching	Task switching (L2)	Task switching (L1)
	Cues for naming in L1:	Cues for naming:	Cues for naming:
	circle, diamond	heart, square	heart, square
	Cues for naming in L2:	Cues for categorizing:	Cues for categorizing:
	triangle, star	cross, hexagon	cross, hexagon
7	Language switching	Task switching (L1)	Task switching (L2)
	Cues for naming in L1:	Cues for naming:	Cues for naming:
	heart, square	circle, diamond	circle, diamond
	Cues for naming in L2:	Cues for categorizing:	Cues for categorizing:
	cross, hexagon	triangle, star	triangle, star
8	Language switching	Task switching (L1)	Task switching (L2)
	Cues for naming in L1:	Cues for naming:	Cues for naming:
	circle, diamond	heart, square	heart, square
	Cues for naming in L2:	Cues for categorizing:	Cues for categorizing:
	triangle, star	cross, hexagon	cross, hexagon

In all three sessions we used the “cuing” paradigm (e.g., Meiran, 1996): the language (in language switching) or the task (in task switching), was specified on each trial by a shape cue (see Fig. 1, top) presented at one of three cue-stimulus intervals (CSIs: 50, 800 or 1175 ms). The language (or task) switched unpredictably and relatively infrequently (switch probability=33%) to discourage participants from anticipating (and preparing for) a switch before they saw the cue (see Introduction). As illustrated in Figure 1, the cue was followed by the presentation of the to-be-named drawing with the cue superimposed for 2000 ms – the time-window during which the response was recorded.

Table 2

Words required as responses in the categorization task (in the task switching sessions) and in the naming task (in the task switching and language switching sessions).

Participant group	Responses in the categorization task (L1/L2) followed by responses in the naming task (L1/L2)			
German	Beruf/Job	Lebensmittel/Food	Körper/Body	Kleider/Clothes
	Bauer/Farmer	Trauben/Grapes	Ohr/Ear	Mantel/Coat
	Lehrerin/Teacher	Ananas/Pineapple	Bein/Leg	Kleid/Dress
	Arzt/Doctor	Käse/Cheese	Gehirn/Brain	Krawatte/Tie
	Kellner/Waiter	Wurst/Sausage	Zähne/Teeth	Gürtel/Belt
French	Métier/Job	Nourriture/Food	Corps/Body	Vêtements/Clothes
	Infirmière/Nurse	Gâteau/Cake	Oreille/Ear	Manteau/Coat
	Facteur/Postman	Raisin/Grapes	Dents/Teeth	Pantalon/Trousers
	Professeur/Teacher	Ail/Garlic	Bra/Arm	Robe/Dress
	Religieuse/Nun	Pomme/Apple	Cou/Neck	Cravate/Tie
Spanish	Oficio/Job	Comida/Food	Cuerpo/Body	Ropa/Clothes
	Enfermera/Nurse	Tarta/Cake	Oreja/Ear	Chaqueta/Coat
	Cartero/Postman	Uva/Grapes	Dientes/Teeth	Pantalones/Trousers
	Profesora/Teacher	Queso/Cheese	Brazo/Arm	Vestido/Dress
	Monja/Nun	Manzana/Apple	Cuello/Neck	Corbata/Tie

The task switch cost has been shown to be sensitive to the interval between the response and the onset of the following stimulus (Meiran, 1996). To unconfound the effects of preparation (CSI) from such effects of the response-stimulus interval we: (1) preceded the cue by a blank screen, whose duration of 2150/1400/1025 ms was inversely dependent on the CSI, to ensure a constant interval (2200 ms) between the end of the response recording window on the previous trial and the onset of the current stimulus; (2) randomized the CSI over trials. Because cue repetition can substantially inflate the task (and language) switch cost (see Introduction), the cue was never repeated from one trial to the next – hence we used 2 cues per task (and per language). Furthermore, to avoid any carryover effects across sessions, different sets of cues were used for language switching and for task switching for each participant. This resulted in a total of 8 cues – white shapes (mean dimensions: 8 x 7.6 mm, SDs: 0.71 mm and 0.35 mm) presented inside a grey square (see Fig. 1, upper panel). The cues were divided into 2 sets – one for language switching and one for task switching – counterbalanced in conjunction with the above-mentioned counterbalancing of the order of sessions (see Table 1).

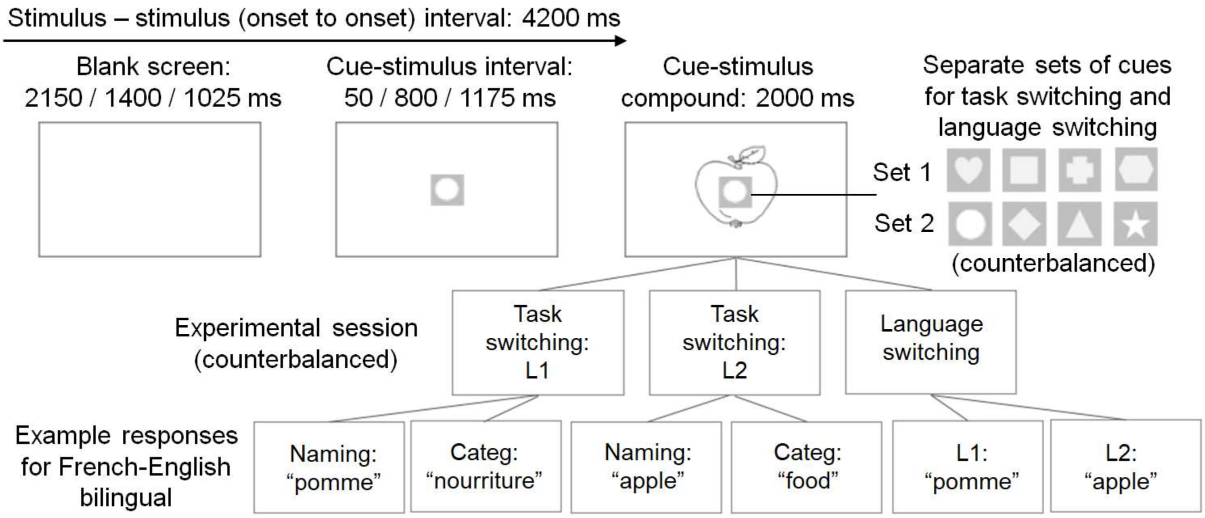

Figure 1. Illustration of the paradigm: the time-course of a trial (top), the cues (top right) and (below) responses as a function of experimental session and language (see Table 1 for the combined counterbalancing of cue sets and sessions over participants).

Every experimental session consisted of 8 blocks each consisting of 72 analyzed trials plus a start-up trial unclassifiable as switch vs. repeat and thus excluded from the analysis. The 576 analyzed trials were uniquely randomly sequenced for every participant and testing session subject to the constraint that in each testing session each of the 16 drawings occurred once on a switch trial and twice on repeat trials for each of 12 combinations of cue x CSI x language/ task (36 times in total), so the contribution of different stimuli (and different responses) to each cell of the experimental design was equal. For the start-up (non-analyzed) trial of every block, the stimulus was selected randomly from the same set of 16 drawings.

Procedure and apparatus

The participant sat in a soundproof cubicle in front of a 19" TFT monitor, wearing a Sennheiser headset with an integrated microphone. The experimental session began with a

10-minute practice session during which the experimenter stood beside the participant in order to provide explanations and correct errors. To familiarise participants with the images and specify the words that had to be used to name the images, each image was presented using E-Prime 1 (Psychology Software Tools, Pittsburgh, PA) with the corresponding word below, which they had to read aloud. Participants self-paced the presentation by pressing the spacebar to continue to the next item. Subsequently, participants had to name each image presented without the word twice – first in a self-paced presentation, then, as in the main part of the experiment, with a 2000 ms response deadline. In language switching sessions this was done in L1 and then repeated in L2. In task switching, participants first named and then categorized the stimuli (in L1 or L2 depending on the language of the session). Participants were then shown the cues that specified either L1 or L2 (in language switching), or naming or categorizing (in task switching). Subsequently, they practiced language (or task) switching, starting with 20 trials all with the longest CSI (1175 ms), where the stimulus was randomly selected among the 16 pictures available (with the constraint that each picture is presented at least once and at most twice), and followed by 32 trials (with each of 16 pictures presented twice in random order) with the three CSIs varying randomly from one trial to another.

Following the practice, the experimenter started the main part of the session conducted using DMDX (Forster & Forster, 2003), then left the testing booth and listened to responses through headphones in the adjacent control room, while monitoring the participant via a camera in the testing booth and a computer in the control room. Following each 73-trial block, a message on the screen reminded the participant to use the cues to prepare to name in the relevant appropriate language or perform the relevant task. Each session lasted ~1 hr. After completing the three sessions, participants were debriefed, paid and asked to complete the language proficiency questionnaire.

RESULTS

Overview of analyses

Our primary aim was to compare the effectiveness of preparation in language switching and task switching, whilst ensuring that the two paradigms were contrasted in the same sample of participants, who were presented with the same cues and stimuli while performing the same task and making the same responses in the two paradigms. To this end, we submitted the speech onset latencies (abbreviated henceforth to RTs for "response times") and the % error rates from the language switching session and from the naming task trials from both task switching sessions (in L1 and in L2) to 'omnibus' repeated-measures ANOVAs with the factors swrep (switch vs. repeat), CSI (50, 800, 1175 ms), paradigm (language switching, task switching) and language (L1, L2). The omnibus ANOVAs were followed-up (when necessary) by ANOVAs that excluded one or more of the above factors; factors whose exclusion is not explicitly stated remained part of the follow-up ANOVAs. For completeness, the categorization task trials (from the task switching sessions) were also analyzed and the outcomes are presented in the Appendix.

We excluded from all analyses filler trials at the beginning of every testing block (unclassifiable as switches vs. repetitions) and trials following errors (most of which also cannot be confidently classified as switches vs. repetitions). Trials on which the participant used a picture name different from that introduced in the practice phase (but which was nevertheless semantically appropriate) were included in the analyses provided the participant used that word consistently (at least twice) in response to the respective image; otherwise they were excluded. In addition, we also excluded from RT analyses trials containing errors, trials where the speech onset could not be determined because it followed the end of the response recording window, and trials on which the speech onset was immediately preceded (and hence likely delayed) by coughing, sneezing, yawning, or other non-speech vocalization. The Huynh-Feldt correction for the violation of sphericity was applied where appropriate, and corrected *p* values are reported, but the degrees of freedom are reported uncorrected. Estimates of effect sizes (η^2_p) are reported in all analyses. In within-subject comparisons (such as those employed here) the variability within individual conditions is uninformative with respect to the variability of the contrasts of interest (i.e., the variability within the switch condition and within the repeat condition tells one nothing about the variability of the switch cost). Thus, in tables and figures the descriptor of variability of interest is the SE of the mean switch cost, rather than the SE of the individual means – we therefore provide the former, but not the latter.

The effect of preparation on the language switch cost and the task switch cost

The omnibus RT analysis revealed statistically significant effects of CSI, $F(2,94)=642.77$, $p<.001$, $\eta^2_p=.932$, reflecting shorter RTs as the CSI increased (see Fig. 2), and paradigm, $F(1,47)=12.02$, $p=.001$, $\eta^2_p=.204$, reflecting shorter RTs in the task switching sessions (922 ms) than in the language switching session (969 ms). There was also a significant main effect of swrep, $F(1,47)=206.78$, $p<.001$, $\eta^2_p=.815$, reflecting the overall performance switch cost (repeat, 917 ms; switch, 974 ms). The switch cost reduced reliably with preparation (swrep x CSI interaction, $F(2,94)=22.16$, $p<.001$, $\eta^2_p=.320$), but this RISC effect was different in the two paradigms, as indicated by the significant swrep x CSI x paradigm interaction, $F(2,94)=6.71$, $p=.002$, $\eta^2_p=.125$. As Figure 2 shows, the early portion of the RISC function was steeper for the task switching paradigm. This was confirmed in follow-up analyses that excluded either the longest or the shortest CSI. The key interaction between swrep, CSI and paradigm was significant in the ANOVA that included CSIs 50 ms and 800 ms, $F(1,47)=10.37$, $p=.002$, $\eta^2_p=.181$, but not the ANOVA that included CSIs 800 ms and 1175 ms, $F(1,47)=1.42$, n.s.

Separate analyses by paradigm showed that for task switching the swrep x CSI interaction was highly significant in the analysis that included all CSIs, $F(2,94)=27.48$, $p<.001$, $\eta^2_p=.369$. Analyses of the task switching paradigm including pairs of CSIs found that increasing the CSI from 50 ms to 800 ms had a large (and highly significant, $F(1,47)=48.11$, $p<.001$, $\eta^2_p=.506$) effect on the switch cost, bringing it to its minimum at CSI=800 ms (see Fig. 2, right panel); the subsequent small increase in the switch cost (from CSI=800 ms to CSI=1175 ms) did not approach significance (F<1). In contrast, the language switch cost reduced modestly and non-significantly from 50 ms to 800 ms, $F(1,47)=2.31$, $p=.136$, $\eta^2_p=.047$ – however, this reduction continued at the same rate beyond CSI=800 ms, hence when all three CSIs were included, the swrep x CSI interaction was significant, $F(2,94)=3.41$, $p=.044$, $\eta^2_p=.068$, and so was the linear component of this interaction, $F(1,47)=7.06$, $p=.011$, $\eta^2_p=.131$. In the language switching paradigm, there was no substantial or statistically detectable asymmetry of switch costs (L1 switch cost, 51 ms; L2 switch cost, 61 ms; language x swrep interaction, $F(1,47)=1.15$, $p=.29$, $\eta^2_p=.024$), or modulation of this asymmetry by CSI (language x swrep x CSI, $F(2,94)=0.98$, $p=.37$, $\eta^2_p=.020$), or indeed an overall difference in response latencies between L1 (967 ms) and L2 (972 ms): main effect of language, $F(1,47)=0.443$, $p=.51$, $\eta^2_p=.009$.

At its minimum, the switch cost was not significantly different between paradigms (paradigm x swrep, $F(1,47)=1.85$, $p<.18$, $\eta^2_p=.038$), and it was significant for both paradigms (task switching when CSI=800 ms, $F(1,47)=29.11$, $p<.001$, $\eta^2_p=.382$; language switching when CSI=1175 ms, $F(1,47)=73.84$, $p<.001$, $\eta^2_p=.611$). The switch cost was however different between paradigms at is maximum (CSI=50 ms) as indicated by the paradigm x swrep interaction for this CSI, $F(1,47)=7.49$, $p<.009$, $\eta^2_p=.137$. The smaller language switch cost when CSI=50 ms (see Fig. 2, upper panel) was primarily due to repeat trials, where RT was significantly longer for language repetitions than for task repetitions, $F(1,47)=9.97$, $p<.003$, $\eta^2_p=.175$, whereas language switches and task switches did not differ significantly for this (shortest) CSI, $F(1,47)=2.84$, $p=.1$, $\eta^2_p=.057$. Conversely, the steeper RISC effect observed for task switching than for language switching from CSI=50 ms to CSI=800 ms is attributable almost exclusively to the switch trials, for which there was a greater improvement over CSIs for the task switching paradigm than for the language switching paradigm (see Fig. 2); for repetitions the benefit of increasing the CSI was similar in the two paradigms. Indeed, in separate ANOVAs for switches and repetitions including only CSIs 50 ms & 800 ms, the paradigm x CSI interaction was significant for the switch trials, $F(2,94)=5.73$, $p=.007$, $\eta^2_p=.109$, but not for the repeat trials, $F<1$.

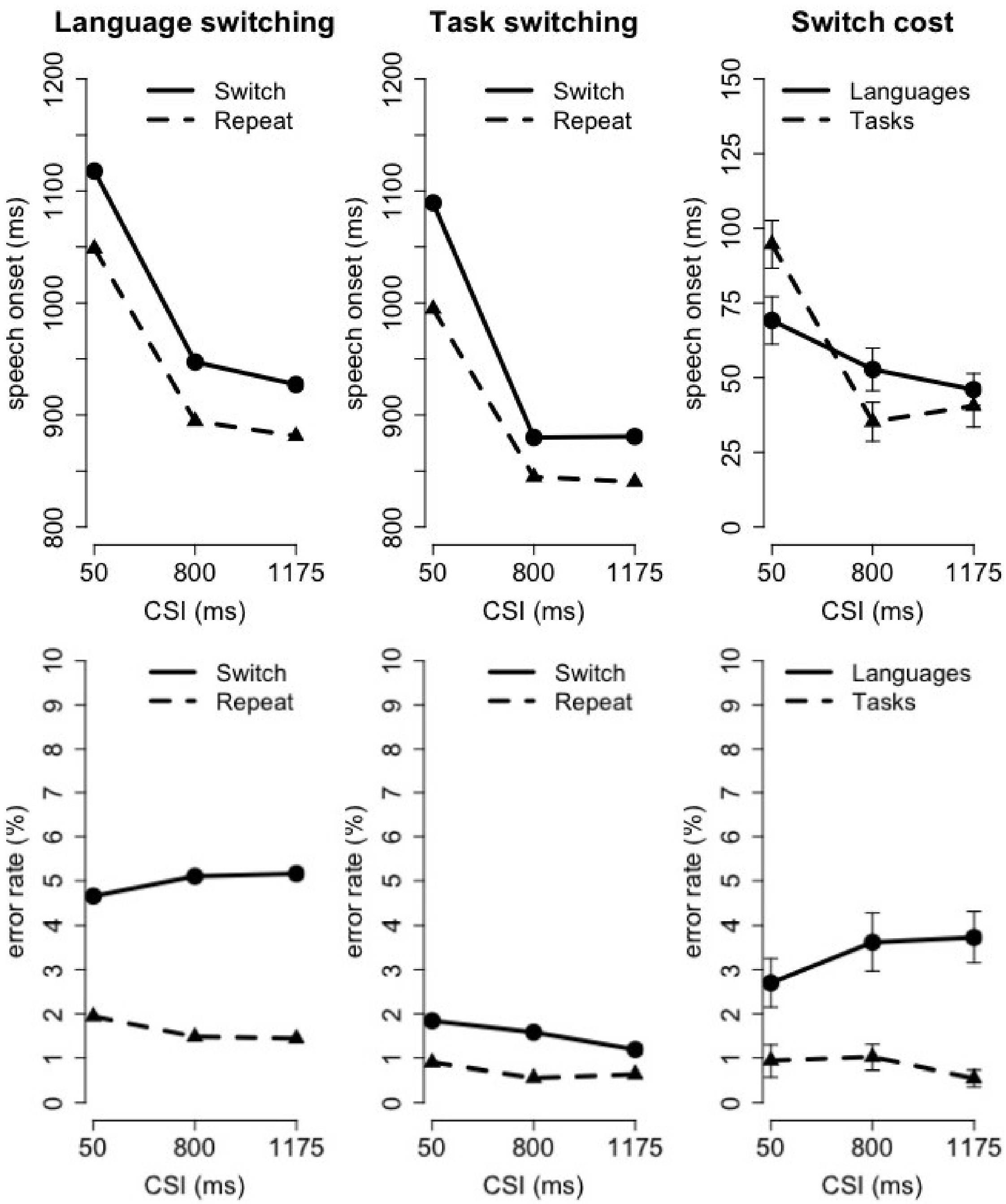

Figure 2. Response onset times (upper panel) and rates of specific ("wrong task" or "wrong language", see text) errors (lower panel) as a function of switch vs. repeat and paradigm (language switching vs. task switching).

Because the testing for the language switching paradigm was conducted in one session, whereas for the task switching paradigm it was conducted over two sessions, one may argue that, although the overall number of naming trials in each language was equivalent for the two paradigms, the split over two sessions may have resulted in a more optimal distribution of practice in task switching, thus leading to a steeper RISC effect. We examined this possibility in two analyses. First, we subjected the naming task data from the task switching sessions to an ANOVA, where the factor language (L1 vs. L2) was replaced with session order (1st vs. 2nd session). The interaction between swrep, CSI and session order did not approach statistical significance, $F(2,94)=0.61$, $p=.548$, $\eta^2_p=.013$, providing no evidence that the small RISC differences between sessions (switch costs in the order of increasing CSI for 1st session: 89 ms, 36 ms, 34 ms; 2nd session: 100 ms, 35 ms, 47 ms) were material. Second, we ran the omnibus ANOVA using the response latencies from the language switching session and only the 1st task switching session[3]. The key interaction between paradigm, swrep and CSI was statistically significant, $F(2,94)=3.83$, $p=.025$, $\eta^2_p=.075$. Thus, we found no evidence that conducting the task switching testing over two sessions and the language switching testing in a single session can explain the differential effect of preparation on switch cost in the two paradigms.

In our error analyses, we separated "specific" errors – naming in the wrong language (in the language-switching session) or categorizing instead of naming (in the task-switching sessions) – and "non-specific" errors – using a semantically inappropriate word and disfluent responses (e.g., pausing before completing the utterance). As shown in Figure 2, the analysis of "specific" errors revealed a significant switch cost [switches, 3.25%; repetitions, 1.16%; main effect of swrep, $F(1,47)=46.48$, $p<.001$, $\eta^2_p=.497$], which was not modulated significantly by preparation (swrep x CSI interaction, $F=1$). There were overall more "wrong language" errors than "wrong task" errors (3.29% vs. 1.12%; main effect of paradigm, $F(1,47)=49.03$, $p<.001$, $\eta^2_p=.511$), and, a substantially larger language switch cost than task

[3] Since for half of the participants this sessions was in L1 and for the other half it was in L2 (and because language did not interact in the above analyses with factors swrep and CSI) we included all the participants' 1st task switching session in the analysis irrespective of language; for the language switching paradigm the data were averaged over language. Hence, this analysis contained the factors *paradigm*, *swrep* and *CSI*, but not the factor *language*.

switch cost for the specific errors [3.35 ± 0.48% vs. 0.84 ± 0.2%; swrep x paradigm interaction, $F(1,47)=36.75$, $p<.001$, $\eta^2_p=.439$].

The analysis of "non-specific" errors found a small switch cost which did not reach significance [switches, 0.84%; repetitions, 0.68%; main effect of swrep, $F(1,47)=3.58$, $p=.065$, $\eta^2_p=.071$], and which was largest in the intermediate CSI [switch cost: CSI=50 ms, -0.03 ± 0.16%; CSI=800 ms, 0.55 ± 0.17%; CSI=1175 ms, -0.03 ± 0.11%; swrep x CSI interaction, $F(2,94)=5.30$, $p=.007$, $\eta^2_p=.101$]. To examine the possibility of L1 vs. L2 switch cost asymmetries, we also run ANOVAs on the "specific" and "non-specific" errors for the language switching paradigm only. There was no sign for either error type of a statistically detectable switch costs asymmetry, or of its interaction with CSI (all *Fs*<2, *ps*>0.15).

The effect of associative history on the language switch cost and the task switch cost

To determine whether the associative bindings formed between the stimulus and the language, or between the stimulus and the task-set, contributed to the switch cost, we examined whether the language (or task) on the most recent encounter with the current stimulus was the same or not as the language (or task) required on the current trial. Thus, for each paradigm separately, we subjected RTs and "specific" errors (task confusions or language confusions) to ANOVAs with the factors swrep, previous encounter, CSI and language. The trial inclusion criteria were the same as for the analyses in the preceding section.

The most recent encounter influenced the overall performance (see Table 3, upper half). RTs were shorter and the errors less likely when (in the language switching session) the picture was most recently named in the same language as currently required than in the other language [main effect of previous encounter, RT, $F(1,47)=96.83$, $p<.001$, $\eta^2_p=.673$, errors, $F(1,47)=16.6$, $p<.001$, $\eta^2_p=.261$], or when (in the task switching sessions) the picture occurred most recently in the context of the same task as the current (naming) task than in the context of the other (categorization) task [main effect of previous encounter, RT, $F(1,47)=132.36$, $p<.001$, $\eta^2_p=.738$; errors, $F(1,47)=9.19$, $p=.004$, $\eta^2_p=.164$].

The key question is whether associative history also contributed to the switch cost: was the switch cost larger when the previous encounter with the stimulus was in the context of the other language or task? The short answer is: 'no'. For task switching, the swrep x previous encounter interaction did not approach significance for either the RT or the error rate, both Fs<1 (other interactions involving both of these factors were also non-significant); the three-way interaction involving these factors and CSI also did not approach significance.

For the language switching condition, the above interaction did not approach significance for the error rate, $F<1$, whereas for the RT, contrary to the associative account of the switch cost (see Introduction), the switch cost was in fact larger when the language of the previous encounter was the same as the currently relevant language (66 ms) than when it was different (40 ms), swrep x previous encounter, $F(1,47)=7.53$, $p=.009$, $\eta^2_p=.138$. Mean RTs show that this was due to shorter repeat RTs (rather than longer switch RTs) in the "same language" condition compared to the "different language" condition (see Table 3), suggesting that language repetition trials may have benefited disproportionately from response priming due to some immediate or near immediate repetitions of the same stimuli and therefore responses.

Table 3

The effects of the language, or the task, in whose context the current stimulus was last encountered, as a function of switch vs. repeat and lags included in the analysis.

Lag between stimulus repetitions	Switch vs. repeat	RT (ms)				Language/task errors (%)			
		Language on previous encounter		Task on previous encounter		Language on previous encounter		Task on previous encounter	
		Same	Dif	Same	Dif	Same	Dif	Same	Dif
Any lag	Switch	979	1014	925	970	4.25	5.56	1.31	1.65
	Repeat	913	974	871	919	1.1	2.17	0.44	0.97
	Switch cost ±	66	40	54	51	3.15	3.39	0.87	0.68
	SE	7.37	5.87	6.81	5.32	0.55	0.55	0.26	0.22
Lag > 4	Switch	981	1017	927	966	4.55	6.25	1.41	1.68
	Repeat	925	972	876	918	1.1	2.03	0.48	0.75
	Switch cost ±	57	45	51	47	3.45	4.22	0.93	0.93
	SE	8.07	6.52	6.57	6.59	0.61	0.61	0.28	0.25

To determine whether response priming may have obscured the effect predicted by the "associative account" of switch cost, we re-ran the analyses for both paradigms, excluding trials for which there were fewer than 5 trials since the last encounter with the same stimulus. As above, these analyses revealed robust effects of associative history on the overall performance (see Table 3, lower half) as reflected by the significant main effect of previous encounter [language switching RT, $F(1,47)=64.21$, $p<.001$, $\eta^2_p=.577$; language switching errors, $F(1,47)=14.97$, $p<.001$, $\eta^2_p=.242$; task switching RT, $F(1,47)=72.53$, $p<.001$, $\eta^2_p=.607$; task switching errors, $F(1,47)=3.24$, $p=.078$, $\eta^2_p=.064$]. Crucially, there were no significant swrep x previous encounter, or swrep x previous encounter x CSI interactions for either paradigm, for either of the two measures (all *F*s<1.75, ns.). Of the four analyses (RTs and errors for each of two paradigms) only in one (language errors) there was a numerical tendency for a larger switch cost in the "previous encounter different" condition – but language RTs showed the opposite numerical trend (see Table 3).

To sum up, our analyses indicate that the context of the previous encounter of the stimulus clearly influenced the overall performance. However, it did not influence the switch cost, at least not in the direction predicted by the associative account of the switch cost: we did not find switch trials to be more susceptible to such associative effects than repeat trials.

The correlation between the language switch cost and the task switch cost

The within-participants design and relatively substantial participant sample enabled us to examine the correlation between the switch costs in the two paradigms. To our knowledge only one other study (Declerck et al., 2017) has reported this correlation for conditions with matched tasks and materials across the two paradigms. The presence of two task switching sessions (in L1 and L2) in our study offers us the additional possibility of comparing the magnitude of the correlation between the language switch cost and the task switch cost with the magnitude of the correlation between two task-switching sessions (cf., Timmer et al., 2018).

We found a moderate, statistically significant, correlation between the RT language switch cost and the RT task switch cost (using the data for the task used in both paradigms – naming), Pearson's $r(48)=.36$, $p=.012$ (see Fig. 3, left panel). This correlation is of comparable magnitude to the correlation between the RT switch costs obtained in the two task switching sessions. Since the temporal order of the different sessions was balanced over participants, we correlated the two task switching sessions (1) by language (L1 session vs. L2 session), ignoring temporal order, $r(48)=.39$, $p=.006$ (see Fig. 3, middle panel), and (2) by temporal (1st vs. 2nd session), ignoring language, $r(48)=.40$, $p=.005$ (see Fig. 3, right panel). None of the three correlations above was reduced when controlling for the overall RT in the respective conditions by means of partial correlations between: language switching and task switching, $r(45)=.36$, $p=.013$; task switching in L1 and task switching in L2, $r(45)=.43$, $p=.002$; task switching in the 1st session and the 2nd session, $r(45)=.46$, $p=.001$.

We also conducted the same correlations for the specific ("wrong task" or "wrong language") errors. There was a substantial, and significant, positive correlation between the language switch cost and the task switch cost, $r(48)=.54$, $p<.001$. However, the very low correlations between the two task switching sessions conducted by either language, or temporal order (*r*(48)=.08 and *r*(48)=.12, both ns.), raise doubts regarding the reliability of the task switch error cost, which was likely subject to a "floor effect" (in many participants/conditions the error task switch cost was null). Hence, the high correlation between the error switch costs in the two paradigm paradigms cannot be treated as conclusive.

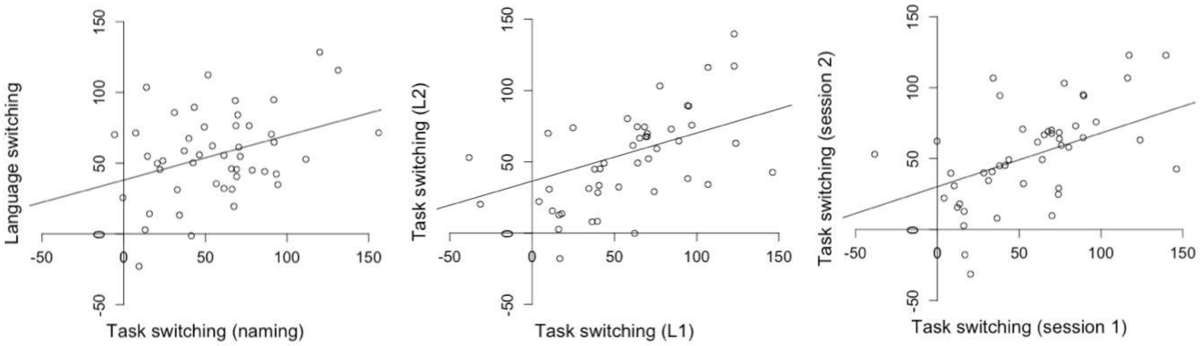

Figure 3. Scatterplots illustrating the correlations between the language switching session and the task switching session (left), and between the two task switching sessions (middle and right).

GENERAL DISCUSSION

The last two decades or so have seen unprecedented synergy between research on bilingualism and on the cognitive control of attention and performance. This synergy has been facilitated by the development of analogous task switching and language switching paradigms which have since been applied extensively to study task-set control and the selection of language for production. The approach which has dominated such investigations thus far has been that of determining the overlap in cognitive processes involved in task switching and language switching by documenting in one domain phenomena previously observed in the other, correlating the magnitude of analogous empirical phenomena in the two domains, etc. (see Introduction).

The current study adopts a somewhat different approach – of comparing the effectiveness of intentional ("top-down", endogenous) control in the two domains. Task switching research has indicated that possibly the "purest" performance measure of intentional control of task-set is the robust reduction in the task switch cost with opportunity for preparation – the RISC effect (see Introduction). Indeed, both empirical data and computational modelling have shown that the RISC effect in task switching cannot be explained away by retrieval of overlearned cue-stimulus compounds in the absence of task-set representations (e.g., Forrest, Monsell, & McLaren, 2014). Thus, we set out to compare the effectiveness of preparing to switch to a linguistic task (picture naming) from another linguistic task (generating a verbal category label) with the effectiveness of preparing to switch languages in a naming task. We matched for all other factors (cues, stimuli, responses, languages, block and trial sequences, and within-trial structure). As reviewed in the Introduction, the RISC effect has been variable in the language switching literature. Perhaps more importantly, recent efforts to control for critical confounds have resulted in little to no reduction in switch cost with preparation, suggesting that preparing a language for production may be less effective than preparing components of task-set. The current study put this conjecture to a methodologically rigorous and well-powered test.

The effectiveness of preparing for a language switch vs. a task switch

Our results show that, although for naming latencies preparation reduced significantly both the task switch cost and the language switch cost, this reduction was far from equivalent in the two paradigms. The early portion of the RISC function was considerably steeper in task switching than in language switching (see Fig. 2, top right panel). Indeed, increasing the CSI from 50 ms to 800 ms resulted in a 63% reduction in the task switch cost and only a 24% reduction in the language switch cost. This difference cannot be explained away simply in terms of a greater task switch cost at CSI = 50 ms, and, hence, greater room for reduction. The difference in switch cost between the paradigms at the shortest CSI was due primarily to repetition trials (language repetitions resulted in longer RTs than task repetitions at CSI = 50 ms, but switch RTs were similar for the two paradigms at this CSI), whereas the difference in RISC slopes from the shortest to the intermediate CSI is primarily attributable to switch trials (see Results and Fig. 2). The task switch cost seemed to reach its asymptote at ~800 ms (it did not reduce with a further increase in the CSI from 800 ms to 1175), whereas the language switch cost continued to reduce at the same rate up to the longest CSI. It is conceivable that a further increase in CSI (>1175 ms) may have resulted in further reduction of the language switch cost. If so, one requires more (longer) preparation for a language switch than for a task switch to achieve a comparable performance benefit.

It is important to stress that the steep RISC function we have observed in the task-switching condition is entirely consistent with RISC effects previously reported in the task-switching literature. Indeed, statistically robust RISC effects of a comparable magnitude have been found for switching between: different semantic classifications of visual words (e.g., Van't Wout et al., 2015); a semantic classification and a phonological classifications of words (e.g., Monsell & Mizon, 2006); identification of colours and identification of canonical geometric shapes (Lavric et al., 2008; Monsell & Mizon, 2006); classification of colours and classification of single letters (Elchlepp, Best, Lavric, & Monsell, 2017); different perceptual classifications of the same set of line drawings (Van't Wout et al., 2015). The nearest to the current picture naming task seems the paradigm used by Van't Wout and colleagues' (2013, Exp. 1), in which participants switched between identifying two sets of pictures (each picture mapped to a computer key). In that experiment increasing the CSI from 100 ms to 1300 ms resulted in a (highly significant) 55% RISC effect. We focused here on studies by our group, partly because we have been especially interested in investigating the RISC effect, and partly because some critical parameters in those studies were the same as they are here (in particular, no immediate cue repetitions and a lower probability of a switch than of a repetition). However, large RISC effects are by no means confined to our studies (c.f., Verbruggen et al., 2007). Moreover, substantial (and statistically robust) RISC effects have also been observed for switches between the visual and auditory perceptual modalities (Lukas et al., 2010) and between voices to listen to in a multitalker compound (Monsell et al., 2019). Hence, of the two divergent RISC functions (steep in our task switching condition, shallow in the language switching condition), the very slim RISC effect in language switching (here and elsewhere, see Introduction) is at odds with the general trend (over different paradigms) for a robust effect of preparation on the switch cost. Why might this be? We discuss two kinds of factor likely at play.

First, as we mentioned in the Introduction, language switching is likely to be associated with a high level of response conflict. Due to extensive experience/exposure, the semantic representation of an object perceived by the bilingual is likely to activate the corresponding phonological representations in both languages, resulting in interference. On a trial where language must change such interference would be especially strong, because the non-target language was recently used. Hence, any preparatory control "bias" in favor of the target language would need to be very strong and sustained in order for it to counteract the imminent interference following stimulus onset. In contrast, during a naming task trial in the task switching paradigm, it is unlikely that the participant experienced the same degree of interference from the categorization task-set, whose S-R rules were only learned and practiced during the experiment. Hence the preparatory bias may not need to be as robust to effectively counteract the interference encountered once the stimulus is perceived. There is some support in our results for the notion that the interference from the non-target language is greater than the interference from the non-target task. First, the “wrong language” error switch cost was much larger than the “wrong task” switch cost (see Fig. 2, lower panels). Second, even on repetition trials, performance was visibly worse in the language switching paradigm than in the task switching paradigm (see Fig. 2), despite the complete equivalence between paradigms with regard to stimuli, responses, trial timing, and parameters of trial sequences (see Method). Third, the idea that the effectiveness of preparation depends on interference leads to an intriguing prediction regarding our pair of tasks, where interference is likely asymmetric. Following perceptual encoding of the stimulus on a categorization trial the (overlearned) responses from the naming task are likely to elicit strong interference, but the reverse interference (on a naming trial) is likely less strong. Categorization performance data (see Appendix) indeed confirm that this is the “weaker” (less practiced, less habitual) of the two tasks. If preparation needs to be more sustained to be effective when interference is strong, the RISC function should be shallower in the task where more interference is experienced. This was indeed the case for the categorization RTs – where the RISC effect was shallower that in the naming task (see Appendix). This pattern parallels the already discussed difference in the RISC functions between language switching and the (naming data from) task switching sessions. In the task switching literature, the comparative effectiveness of preparation in an asymmetric task pair has received very little scrutiny. To our knowledge only one experiment has looked into this (Yeung & Monsell, 2003, Experiment 4) – but it is reassuring that there, as in our task switching data, the RISC effect was steeper in the “stronger” (more recently practiced) of the two tasks.

The second factor that may explain the differential effectiveness of preparation in reducing language switch costs vs. task switch costs is that a task switch tends to involve either a change in the relevant perceptual attribute (e.g., colour vs. form, Elchlepp, et al., 2017), or a change in the semantics of S-R mappings, e.g., classifying a number by parity or by magnitude, or both. In our study, task switches did not require changes in perceptual selection, but the responses in the two tasks were semantically different (item labels vs. category labels). The change from the S-R rules of one task to the S-R rules of the other task likely involves activation of the relevant semantic information. If performed in advance of the stimulus, such semantic processing can serve as an extra bias in favor of the relevant set of responses (and against the irrelevant response set). When a task switch requires a change in the relevant perceptual attribute, preparatory attentional selection (e.g., Longman, Lavric, & Monsell, 2013; Mayr, Kuhns, & Rieter, 2013) may also help subsequent response selection – by reducing the encoding of the irrelevant perceptual attribute, and thus decreasing its potency in activating one of the irrelevant responses. A language switch involves neither a (substantive) change in the meaning of the responses, nor a change in the relevant perceptual attribute/dimension, hence preparation for a language switch cannot benefit from such anticipatory biases. One could, of course, envisage the possibility of a "global" top-down bias in the output (production) lexicon towards items in one language and/or away from items in other languages (this would seem to require a superordinate "language" representation connected to all items in a given language)[4]. If this form of endogenous language control in production indeed exists, there is no reason for it not to be exerted in anticipation of the stimulus increasing the baseline activation of lexical items in the target language relative to items in the non-target language(s). But, the present results suggests this either does not happen or it happens to a limited extent – perhaps because such widespread/diffuse biasing could result in extra lexical competition in the target language (especially if one factors in some noise); it would also be extremely energetically costly for the brain[5]. Thus, during a language switch there may be fewer components of the task-set that are "(re)configurable" in advance of the stimulus. We return to both factors considered above in the concluding part of the discussion.

Can stimulus-language associations (at least in part) explain the language switch cost?

We were also interested in the contribution of stimulus associative history to the language switch cost. In particular, in task switching it has been found that the switch cost was larger for stimuli previously encountered in the context of the currently irrelevant task than for stimuli previously encountered both in the context of the currently relevant task and in the context of the currently irrelevant task (e.g., Waszak et al., 2003; see Introduction). This empirical phenomenon has been the basis for the notion that stimuli form associative links ("bindings") with task-set(s), which subsequently influence performance and in particular the switch cost. For example, if a stimulus has a stronger association with the currently irrelevant

[4] We thank one of the reviewers for raising this possibility.

[5] A related issue is that language switching typically involves relatively large sets of S-R mappings (in the current experiment each set comprised 16), which makes it unlikely that all (or most of) the S-R mappings for the target language are activated during preparation.

task-set than with the currently relevant task-set (e.g., because it was previously only, or more recently, encountered in the context of the former), its presentation will reactivate the irrelevant task-set leading to increased task-set competition. To determine whether this "associative account" would also hold for language switching, we used the most recent encounter with the stimulus as a proxy for the relative strength of the association between the stimulus and the two languages (in the language switching session) or task-sets (in task switching sessions).

Our results have indeed revealed robust associative effects in both paradigms, with better performance when the stimulus' most recent encounter was in the same language (or task) as that required on the current trial (see Table 3). This effect was statistically highly significant in nearly all analyses of RT and error rate, indicating that our measure of associative history had more than adequate sensitivity. However, crucially, there is little indication of even a numerical trend towards a larger switch cost when the language (or task) on the previous encounter was different from the current target language (or task) vs. when it was the same. Indeed, such a numerical trend was only present for the task errors (where it did not approach significance), whereas for RTs the numerical trends in both task switching and language switching were, if anything, in the opposite direction. These results provide little support for associative bindings as a source of the switch cost as previously proposed in the task switching literature (Waszak et al., 2003; 2004; 2005). They also raise questions about the ubiquity of the previously reported interaction between associative history and the task switch cost. In our data this interaction is absent not only in the naming task, but also in the categorization task (see Appendix). There too, the effect of associative history on the overall performance is robust, yet there is no indication that associative history influenced the switch cost. We also note that there have been other failures to observe the associative history x switch interaction (Koch, Prinz, & Allport, 2005), and that associative history cannot, on its own, explain why stimulus-task-set associations should have a greater effect on switch trials than on repeat trials, hence, one must invoke another factor (e.g., task-set inertia, Allport et al., 1994) to explain why switch trials should be more susceptible to such an effect (cf., Waszak et al., 2005).

Recently, Kleinman and Gollan (2018) have conducted a related "item history" analysis of data from several of their language switching experiments and found that more encounters with a given picture in the same language monotonically improved naming performance (reduced naming latencies), whereas more encounters with a given picture in the other language monotonically hindered performance (increased naming latencies). Not only are those effects consistent with the results reported here – like us, Kleinman and Gollan found no interaction between item history effects and the language switch cost. However, the authors did not interpret their results in terms of associative bindings, favoring instead a "persisting/accumulating inhibition" account, where naming a picture is accompanied by inhibition of the semantically corresponding lemma in the other language – this inhibition persists and accumulates with further instances of naming the same picture. As far as we can tell, item history effects in language switching can be explained by either associative bindings or persisting inhibition or indeed some combination of the two. However, as acknowledged by Kleinman and Gollan, persisting inhibition of the competing response is unlikely to adequately account for the item history effects in task switching, where previous encounters in the context of the currently irrelevant task is detrimental to performance even for "response congruent" stimuli – which require the same response in both tasks (Koch & Allport, 2006). Thus, parsimony favors an associative account of item history effects – it doesn't require different mechanisms for task switching vs. language switching, and within language switching between response priming vs. persisting inhibition. Finally, we note that even if persisting inhibition explains a (substantial) part of the item history effect in our language switching condition, this does not take away from the conclusion we draw from this analysis that, inasmuch as stimulus-language associations form during the experiment, they are not among the factors contributing to the language switch cost.

Does the language switch cost correlate with the task switch cost?

If (some of) the processes that are the sources of these performance costs are shared across domains, one would expect at least a moderate correlation. So far, the evidence on the magnitude of this correlation is mixed. Low correlations have been reported in studies where the language switching and task switching paradigms used different stimuli, responses and tasks. Moderate to high correlations were observed when at least some of these parameters were matched – but so far only in one study (Declerck et al., 2017), which, unfortunately, did not unconfound the switch costs from the cue change/repetition effect. Among Declerck et al.'s experiments, our design is most similar to their Experiment 3. There, as in our experiment, the task switching condition involved switching between picture naming and picture categorization. However, there were also a few non-trivial differences, of which possibly the most relevant ones were that (as already mentioned) they used one cue per language (and per task), their language switching condition contained not only picture naming, but also picture categorization (in separate blocks), they used only 4 picture stimuli (presumably to equate the number of possible responses in the naming and categorization tasks), and their testing was conducted in a single session. Some of these differences would likely account for the difference in the magnitude of the correlation (0.36 in our experiment vs. 0.44 – 0.64 in theirs). For instance, multiple testing sessions in our experiment likely introduced extra variance; the cue change/repetition effect may have inflated the variance shared by the two paradigms in Declerck et al.'s experiment, and so forth. However, albeit somewhat smaller, our significant correlation is broadly consistent with their findings. Together, these results suggest that some of the control processes that enable the cognitive system to change the language for production are also those that enable the change of the semantic and response components of a linguistic task-set (switching from categorization to naming). What may be the overlapping processes? We speculate that it is the top-down biasing of (phonological) lexical selection, which has to be (re)configured both during a language switch and during a change of linguistic task. With regard to the previously reported low (or near-null) correlations between the language switch cost and the cost of switching between non-linguistic tasks (e.g., colour-shape switching), these appear to speak against a general (universal) "shifting" control mechanism (e.g., Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000). Indeed, it is conceivable that the switching of some components of task-set (e.g., perceptual encoding) may work differently to switching of other components (e.g., lexical selection).

Relevant neuroscience evidence and final considerations

In the current study, we set out to investigate and relate language switching performance to task switching performance. Most importantly, we compared the effectiveness of intentional (top-down) control in the two domains by focusing on the preparatory component of intentional control. Our primary finding is that effective preparation takes longer when bilinguals switch between languages for production than when they switch between instructed tasks. We return here to the two potential interpretations we suggested earlier for this result – one in terms of greater exogenous (stimulus-related) interference which preparation has to prevent/counteract, and the other in terms of preparatory semantic processes that may accompany the activation of the relevant set of S-R rules during a task switch, but not (or not as much) during a language switch[6]. In our discussion of these two interpretations we have not thus far considered one key difference between them. The first

[6] We assume that the meanings of the word responses used in the two languages are very similar.

interpretation does not need to assume that preparation for a language switch vs. task switch includes qualitatively different processes, only that similar preparatory processes take longer to be effective for the former relative to the latter. In contrast, qualitatively different preparatory processes are inherent to the second interpretation – which assumes that preparation for a task switch contains a process which is absent in selecting a language for output.

Given this important distinction, evidence from neuroscience may help decide between the two interpretations. If preparation to change tasks involves qualitatively different processes from preparation to change languages, the two kinds of preparation should also be reflected in qualitatively different patterns of brain activity. Conveniently, cognitive electrophysiologists have isolated a robust "signature" of effective preparation for a task switch – an EEG-derived event-related potential switch vs. repeat difference of positive polarity which is maximal over the posterior scalp, hence the terms "posterior positivity" or "parietal positivity" (e.g., Karayanidis, Coltheart, Michie, & Murphy, 2003). Not only is it ubiquitous in the late part of the preparation interval of task-switching EEG studies that examined preparation (for a review, see Karayanidis, Jamadar, Ruge, Phillips, Heathcote, & Forstmann, 2010), it has been shown to predict switching performance both within participants (where it is small or absent on trials with a large switch cost and substantial on trials with a small switch cost, e.g., Karayanidis, Provost, Brown, Paton, & Heathcote, 2011; Lavric et al., 2008) and over participants (participants who show the largest posterior positivity are also those showing the steepest RISC effect, Elchlepp, Lavric, Mizon, & Monsell, 2012; Lavric et al., 2008).

Does preparation for a language switch also elicit the posterior positivity? A recent study from our laboratory (Lavric et al., 2019) showed that it does, and that, as in task switching, its magnitude predicts the effectiveness of switching performance. The above evidence of the switch-related positivity in both paradigms is consistent with qualitatively similar preparatory processes in the two domains. Furthermore, although the relatively low time-resolution of fMRI does not enable unambiguous separation between preparatory (pre-stimulus) and post-stimulus brain activity, a recent fMRI study that compared the activations elicited by the two kinds of switching found more commonality than differentiation (De Baene, Duyck, Brass, & Carreiras, 2015). On these grounds, of the two interpretations of our results we have discussed earlier, we tentatively favor the one that does not need to assume the existence of qualitatively different preparatory processes in language switching and switching between linguistic tasks. According to this interpretation, stimulus-elicited interference tends to be greater in language switching than in task switching – hence, to pre-emptively counteract this interference, preparation for a language switch needs to be more sustained. This account also appeals to us because, as already noted, it naturally explains why (in the task switching data, see Appendix) the RISC function is shallower for the task which can be assumed to be subject to greater interference from the competing task following stimulus onset – categorization (subject to stronger interference from naming than vice-versa).

An important question which remains unanswered is whether the language switch cost in our longest CSI (1175 ms) was asymptotic. Although the RISC from CSI=800 ms to CSI=1175 ms was not statistically significant, the clear (and significant) linear trend over the three CSIs suggests that a further increase in the CSI may well result in further reduction in the language switch cost. Thus, a key topic for future research is to determine whether, given ample time (or opportunity for participants to control the preparation interval, cf. Longman, Lavric, & Monsell, 2017) preparation for a language switch would match the asymptotic effectiveness observed in task switching.

CONTEXT

This research has emerged from efforts to cross-fertilize the empirical and theoretical frameworks of research on task switching and bilingualism. We, and others, have found that when participants are introduced to simple laboratory tasks, even after limited practice, they can very effectively "set themselves up" in advance for a task change (or "switch") – so much so that preparation greatly reduces the performance "switch cost". Moreover, EEG studies have revealed an electrophysiological correlate of preparing to switch tasks. Our recent research has shown that preparing to speak another language was associated with a very similar EEG signature – but the benefit of preparation to language switching performance seemed modest when key confounds were addressed. We therefore set out to determine if "setting oneself up" in advance for a language switch is indeed less effective (more time consuming). In addition to examining preparation, we wanted to test in the bilingual control domain the "associative history" account of the switch cost, which has been influential in task switching. It posits that stimuli form associative bindings with the task-sets (here: languages) in whose context they occur, and that subsequent presentation of a stimulus will retrieve via these associations the relevant or irrelevant task-set (here: language).

REFERENCES

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M Moscovitch (Eds.), *Attention and Performance XV: Conscious and nonconscious information processing* (pp. 421-453). Cambridge, Massachusetts: The MIT Press.

Bialystok, E., & Craik, F. (2015). Cognitive consequences of bilingualism. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 571-585). Cambridge University Press

Bialystok, E., Craik, F., & Freedman, (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, 45, 459-464.

Branzi, F. M., Calabria, M., Boscarino, M. L., & Costa A. (2016). On the overlap between bilingual language control and domain-general executive control. *Acta Psychologica*, 166, 21-30.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9.

Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2012). Qualitative differences between bilingual language control and executive control: evidence from task-switching. *Frontiers in Psychology, 2*, 1-10.

Chabal, S., & Marian, V. (2015). Speakers of different languages process the visual world differently. *Journal of Experimental Psychology: General,* 144, 539-550.

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.

Council of Europe (2001). Common European Framework of Reference for languages: Learning, Teaching, Assessment. Cambridge University Press, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09 00001680459f97

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.

Declerck, M., Ivanova, I., Grainger, J., & Duñabeitia, J. A. (2020). Are similar control processes implemented during single and dual language production? Evidence from switching between speech registers and languages. *Bilingualism: Language and Cognition, 23,* 694-701.

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

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

Elchlepp H., Best, M., Lavric A., & Monsell S. (2017). Shifting attention between visual dimensions as a source of switch costs. *Psychological Science*, 28, 470-481.

Elchlepp, H., Lavric A., Mizon, G. & Monsell, S. (2012). A brain-potential study of preparation for and execution of a task-switch with stimuli that afford only the relevant task. *Human Brain Mapping*, 33, 1137-1154.

Elchlepp, H., Lavric A., & Monsell, S. (2015). A change of task prolongs early processes: evidence from ERPs in lexical tasks. *Journal of Experimental Psychology: General*, 144, 299-325.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Fink, A., & Goldrick, M. (2015). Pervasive benefits of preparation in language switching. *Psychonomic bulletin & review*, *22*, 808-814.

Forrest, C.L.D., Monsell, S., and McLaren I.P.L. (2014). Is performance in task-cuing experiments mediated by task-set selection or associative compound retrieval? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 40, 1002-1024.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116-124.

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

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

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

Heikoop, K., Declerck, M., Los, S., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language switching. *Bilingualism: Language and Cognition*, 19, 921-927.

Hilchey, M., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks? In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 586-613). Cambridge University Press

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, 14, 89, 81.

Karayanidis, F., Coltheart, M., Michie, P.T. & Murphy, K. (2003). Electrophysiological correlates of anticipatory and poststimulus components of taskswitching. *Psychophysiology*, 40, 329-348.

Karayanidis, F., Jamadar, S., Ruge, H., Phillips, N., Heathcote, A., & Forstmann, B. U. (2010). Advance preparation in task-switching: Converging evidence from behavioral, brain activation, and model based approaches. *Frontiers in Psychology*, 1, 25.

Karayanidis, F., Provost, A., Brown, S., Paton, B., & Heathcote, A. (2011). Switch-specific and general preparation map onto different ERP components in a task-switching paradigm. *Psychophysiology*, 559-568.

Kelcha, A. (2013). Language and task switching in Polish-English bilinguals. *Psychology of Language and Communication, 17, 17-36*

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

Kikumoto, A., Hubbard, J., & Mayr, U. (2016). Dynamics of Task-set Carry-Over: Evidence from Eye-Movement Analyses. *Psychonomic Bulletin and Review*, 23, 899-906.

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*(5), 700-714.

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

Koch, I., & Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Memory & Cognition, 34*, 433-444.

Koch, I., Lawo, V., Fels, J., & Vorländer, M. (2011). Switching in the cocktail party – exploring intentional control of auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1140-1147.

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.

Lavric, A., Clapp, A., East, A., Elchlepp, H., & Monsell, S. (2019). Is preparing for a language switch like preparing for a task switch? *Journal of Experimental Psychology: Learning, Memory and Cognition, 45,* 1224-1233.

Lavric, A., Mizon, G.A., & Monsell, S. (2008). Neurophysiological signature of effective anticipatory task-set control: a task-switching investigation. *European Journal of Neuroscience*, 28, 1016-1029.

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.

Longman, C.S., Lavric, A., Monsell, S. (2013). More attention to attention? An eye-tracking investigation of selection of perceptual attributes during a task switch. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1142-1151.

Longman, C. S., Lavric, A., & Monsell, S. (2016). The coupling between spatial attention and other components of task-set: A task-switching investigation. *Quarterly Journal of Experimental Psychology*, 69, 2248-2275.

Longman C. S., Lavric A., & Monsell S. (2017). Self-paced preparation for a task switch eliminates attentional inertia but not the performance switch cost. *Journal of Experimental Psychology: Learning Memory and Cognition*, 43, 862-873.

Longman, C.S., Lavric, A., Munteanu, C. & Monsell, S. (2014) Attentional inertia and delayed orienting of spatial attention in task-switching. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1580-1602.

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.

Lukas, S., Philipp, A. M., & Koch, I. (2010). The role of preparation and cue-modality in crossmodal task switching. *Acta Psychologica*, 134, 318-322.

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General*, 129, 4-26.

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

Mayr, U., Kuhns, D., & Rieter, M. (2013). Eye movements reveal dynamics of task control. *Journal of Experimental Psychology: General*, 142, 489-509.

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1423-1442.

Meiran, N., & Marciano, H. (2002). Limitations in advance task preparation: Switching the relevant stimulus dimension in speeded same-different comparisons. *Memory & Cognition*, 30, 540-550.

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

Miyake, A., Freidman, 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.

Monsell, S. (2015) Task-set control and task switching. In J. Fawcett, E. F. Risko, & A. Kingstone (Eds), *The Handbook of Attention*, Ch, 7 (pp.139-172). MIT Press.

Monsell, S. (2017) *Task-set regulation*. In T. Egner (Ed.), The Wiley Handbook of Cognitive Control, Ch. 2 (pp. 29-49). Wiley-Blackwell.

Monsell, S., Lavric, A., Strivens, A., & Paul, E. (2019). Can we prepare to attend to one of two simultaneous voices? *Journal of Experimental Psychology: Human Perception and Performance*, 45, 966-982.

Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology: Human perception and performance, 32*, 493-516.

Mosca, M., & Clahsen, H. (2016). Examining language switching in bilinguals: The role of preparation time. *Bilingualism: Language and Cognition, 19*, 415-424.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66, 232-258

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

Philipp, A. M. & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching among languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1187-1195.

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.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com.

Rogers, R. D., & Monsell, S. (1995). Costs of predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207-231.

Shaffer, L. H. (1965). Choice reaction with variable S – R mapping. *Journal of Experimental Psychology. 70*, 284 – 288

Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *The American Journal of Psychology, 89*, 669-679.

Stasenko, A., Matt, G. E., & Gollan, T. H. (2017). A relative bilingual advantage in switching with preparation: Nuanced explorations of the proposed association between bilingualism and task switching. *Journal of Experimental Psychology: General, 146,* 1527-1550.

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 89-103.

Timmer, K., Calabria, M., Branzi, F. M., Baus, C., & Costa, A. (2018). On the reliability of switching costs across time and domains. Frontiers in Psychology, 9, 1032.

Titone, D., & Baum, S. (2014). The future of bilingualism research: Insufferably optimistic and replete with new questions. *Applied Psycholinguistics*, 35, 933-942.

Vandierendonck, A., Liefooghe, B., Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and inference control. *Psychological Bulletin, 136,* 601-626.

Van't Wout, F., Lavric, A., & Monsell, S. (2013). Are stimulus-response rules represented phonologically for task-set preparation and maintenance? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1538-1551.

Van't Wout, F., Lavric, A., & Monsell, S. (2015). Is it harder to switch among a larger set of tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41* 363-376.

Verbruggen, F., Liefooghe, B., Vandierendonck, A., & Demanet, J. (2007). Short cue presentations encourage advance task preparation: a recipe to diminish the residual switch cost. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 342-356.

Waszak, F., Hommel, B., & Allport, A. (2003). Task-switching and long- term priming: Role of episodic stimulus-task bindings in task-shift costs. *Cognitive Psychology*, 46, 361–413.

Waszak, F., Hommel, B., & Allport, A. (2004). Semantic generalization of stimulus-task bindings. *Psychonomic Bulletin & Review, 11*, 1027-1033.

Waszak, F., Hommel, B., & Allport, A. (2005). Interaction of task readiness and automatic retrieval in task switching: Negative priming and competitor priming. *Memory & Cognition*, 33, 595–610.

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.

Yeung, N., & Monsell, S. (2003). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance,* 29, 455-469.

APPENDIX: RESULTS FOR THE CATEGORIZATION TASK

Our main primary set of analyses focussed on the task that was matched (common) for the language switching and task switching paradigms – the naming task. However, for completeness, we have also analyzed the performance in the categorization task. Swrep x CSI x Language ANOVAs found a significant switch costs for both RT and errors, as indicated by the main effect of switch for both measures, $F(1,47)=116.81$, $p<.001$, $\eta^2_p=.713$; $F(1,47)=14.52$, $p<.001$, $\eta^2_p=.236$, respectively (see Fig. A1, right panels). While for errors no other effects were significant, for RTs there was also a significant main effect of CSI, reflecting faster responses in the long CSI, $F(2,94)=411.53$, $p<.001$, $\eta^2_p=.897$, and a significant language x swrep interaction, $F(1,47)=7.03$, $p<.011$, $\eta^2_p=.013$, reflecting a somewhat smaller switch cost in L2 (switch, 1113 ms; repeat, 1068 ms) than in L1 (switch, 1131 ms; repeat, 1064 ms).

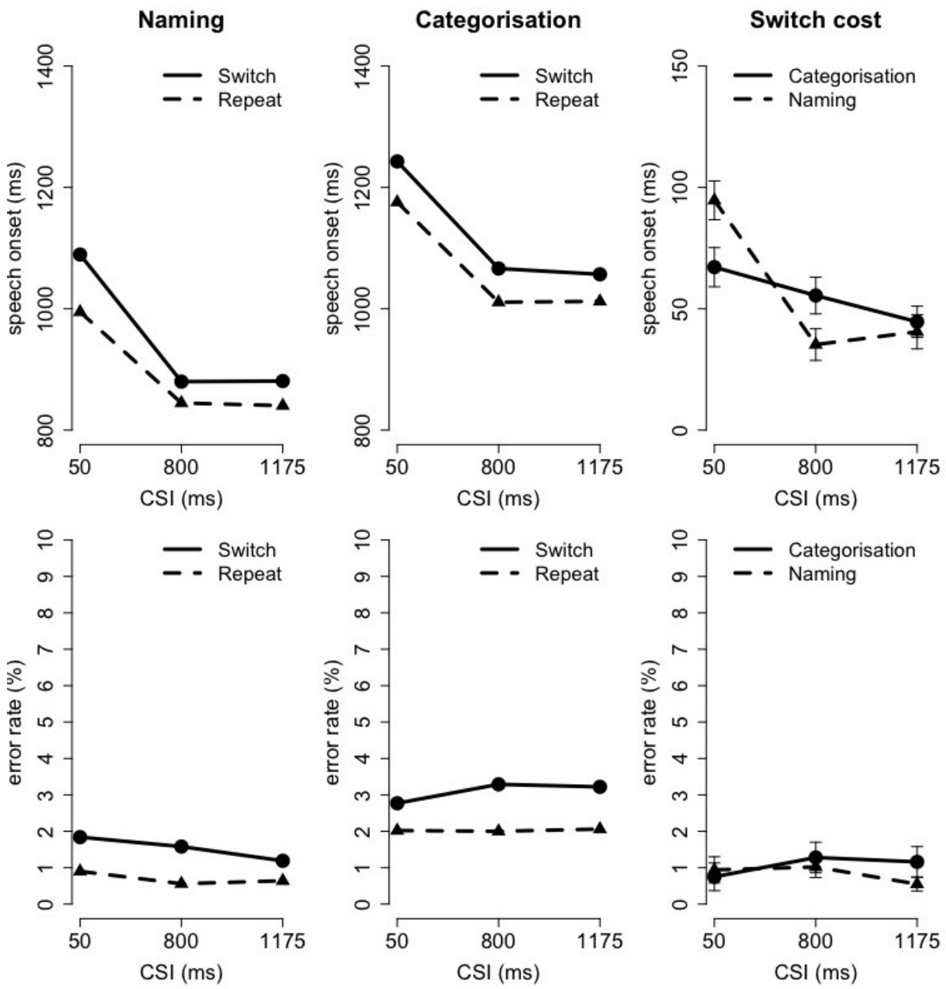

Figure A1. Response onset times (upper panel) and error rates (lower panel) as a function of switch vs. repeat and task. The naming task performance was already presented in Figure 2 (see Results).

The switch cost reduced with preparation for RTs, but this RISC effect was only marginally significant, $F(2,94)=3.03$, $p<.053$, $\eta^2_p=.061$, and appeared less steep than in the naming task (see Fig. A1, right upper panel). This was confirmed in an ANOVA that included both tasks by the significant swrep x task x CSI interaction, $F(2,94)=7.65$, $p<.001$, $\eta^2_p=.14$. This ANOVA also found a significant main effect of task, $F(1,47)=408.63$, $p<.001$, $\eta^2_p=.897$, reflecting substantially longer response latencies in the categorization task (1094 ms) than in the naming task (922 ms). The above-mentioned difference between the RISC effects in the two tasks was primarily driven by the steeper RISC in the naming task between the CSIs of 50 ms and 800 ms, as confirmed by the significant swrep x task x CSI interaction in the ANOVA that excluded the longest CSI, $F(1,47)=14.54$, $p<.001$, $\eta^2_p=.236$, and the non-significance of the same interaction in the ANOVA that excluded the shortest CSI, $F(1,47)=1.54$, n.s. Finally, the shallower reduction in switch cost from CSI=50 ms to CSI=800 ms in the categorization task was due primarily to switch trials, as indicated by separate ANOVAs for switches and repetitions including only these two CSIs – there the task x CSI interaction was significant for switches, $F(1,47)=9.30$, $p=.004$, $\eta^2_p=.165$, but did not reach significance for repeats, $F(1,47)=3.66$, $p=.062$, $\eta^2_p=.072$.

The analysis of the associative history effects in the naming task (see Results) found that associative history influenced performance, but that it did not influence the switch cost. To determine whether this conclusion would also apply to the categorization task, we have conducted the same kind of analysis for this task. As in the main set of analyses, we started with a previous encounter x swrep x CSI x language ANOVA that included all the trials irrespective of the lag since the last encounter with the current stimulus. As for the other (naming) task, associative history robustly influenced categorization performance for both switch and repeat trials (see the upper half of Table A1), as confirmed by the highly significant main effect of previous encounter for both RTs, $F(1,47)=156.16$, $p<.001$, $\eta^2_p=.769$, and errors, $F(1,47)=6.93$, $p=.011$, $\eta^2_p=.128$. Crucially, there is no indication that having encountered the stimulus in the other task (naming) increased the switch cost. Although the errors showed a very small trend in this direction, $F<1$, ns., RTs contained a significant difference in the opposite directions, with a smaller switch cost when the stimulus has been last encountered in the other task (previous encounter x swrep interaction, $F(1,47)=14.5$, $p<.001$, $\eta^2_p=.236$). The inspection of the means for this interaction shows that, as in the language switching data (see Results), the interaction is primarily driven by repeat trials which benefited more than switches from having last encountered the stimulus in the same task, which was likely due to response priming. To minimise the potential effect of response priming, we have (as in the main analysis, see Results) re-run the analysis including only trials where the most recent encounter with the current stimulus was at least 5 trials ago. Here, again, associative history had a clear effect on performance (see the lower half of Table A1), though significantly so only for response onset latencies – these were shorter when the stimulus was last encountered in the categorization task (1089 ms) than when it was last encountered in the naming task (1119 ms), $F(1,47)=46.45$, $p<.001$, $\eta^2_p=.497$. Importantly, as can be seen in Table A1, there was again no indication that associative history increased the switch cost – the previous encounter x swrep interactions did not approach significance for either RTs or errors, $Fs<1$, ns.

Table A1

The effects of the task in whose context the current stimulus was last encountered, as a function of switch vs. repeat and lags included in the analysis.

	RT (ms)				Errors (%)			
	Any lag between stimulus repetitions		Lag > 4 between stimulus repetitions		Any lag between stimulus repetitions		Lag > 4 between stimulus repetitions	
Stimulus previously encountered in	Cat (same) task	Nam (dif) task	Cat (same) task	Nam (dif) task	Cat (same) task	Nam (dif) task	Cat (same) task	Nam (dif) task
Switch	1107	1135	1111	1140	2.67	3.31	2.95	3.08
Repeat	1039	1100	1068	1098	1.71	2.30	1.99	2.24
Switch cost ± SE	68	35	43	42	0.96	1.02	0.96	0.84
	7.70	5.73	7.54	6.58	0.30	0.36	0.37	0.39