Bilingualism and executive attention: evidence from studies of proactive and reactive control

Giacomo Spinelli[1], Samantha F. Goldsmith[2], Stephen J. Lupker[2], J. Bruce Morton[2]

[1] Università degli Studi di Milano-Bicocca

[2] University of Western Ontario

Running Header: Bilingualism and executive attention

Corresponding Author:

Giacomo Spinelli

Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca

Piazza dell'Ateneo Nuovo 1, 20126 Milano (MI), Italy

E-mail: giacomospinelli@hotmail.it

Word count: 15251

Abstract

According to some accounts, the bilingual advantage is most pronounced in the domain of executive attention rather than inhibition and should therefore be more easily detected in conflict adaptation paradigms than in simple interference paradigms. We tested this idea using two conflict adaptation paradigms, one that elicits a list-wide proportion-congruent effect and one that elicits an item-specific proportion effect. In both cases, the relevant finding is that congruency effects are reduced when the proportion of congruent to incongruent items is smaller. These effects are validated measures of proactive and reactive control, respectively, aspects of executive attention known to be associated with individual differences in working-memory capacity. We reasoned that if bilingualism affects executive attention in a similar way as does working-memory capacity, indices of proactive and reactive control should be comparably associated with continuous variation in language status and working-memory capacity. In two experiments, we replicated previous findings that working-memory capacity is associated with variation in congruency effects (suggesting greater reliance on proactive control). In contrast, language status had no consistent association with performance, save for a hint that bilingualism may be associated with greater reliance on reactive control. Thus, the bilingual advantage may exist, but not in proactive control or any other aspects of executive attention that have been proposed thus far.

Keywords: bilingual advantage; proactive control; reactive control; executive attention; working-memory capacity

Bilingualism and executive attention: evidence from studies of proactive and reactive control

It has been long thought that bilinguals are advantaged in cognitive control relative to monolinguals (e.g., Bialystok et al., 2009; Bialystok et al., 2004) owing to regular experience managing two languages. A recent wave of null findings, however, has called the empirical basis of the bilingual advantage into question and led to new ideas about the nature and measurement of language status and its effects on cognitive control (e.g., Duñabeitia et al., 2014; Hilchey & Klein, 2011; Paap & Greenberg, 2013; Paap et al., 2015). For example, in the first meta-analytic review on the topic, Hilchey and Klein (2011) rejected the dominant idea at the time that the bilingual advantage resides in inhibitory processes, the set of processes whereby selection of task-relevant information is made possible by suppression of task-irrelevant information, and proposed instead that the advantage resides in a more general executive monitoring process (but for a revision, see Hilchey et al., 2015).

A more recent model (Bialystok, 2017) links the presumed bilingual advantage to differences in executive attention – processes that support the active maintenance of attention-guiding rules and that vary with individual differences in Working Memory (WM) capacity. This reconceptualization of the bilingual advantage account moves away from traditional explanations based on the concept of inhibition and points to possible parallels between language status and WM capacity-related differences in cognitive control.

Beyond these debates, there also is growing discussion about the measurement of language status itself. One important idea in this regard is that although language status is typically measured categorically, in reality language status varies continuously from individual to individual. One implication is that complex and nuanced relationships between language status and cognition are obscured in studies that use categorical rather than continuous measures of language status (Luk & Bialystok, 2013).

To begin exploring these ideas, Grundy et al. (2017) tested for language status differences on conventional and sequential congruency effects (SCEs) measured by means of the Eriksen flanker task (Eriksen & Eriksen, 1974). In the flanker task, participants identify a centrally presented target (e.g., an arrow) among flanking distractors (e.g., other arrows). The conventional congruency effect refers to the fact that responding is typically slower and less accurate for targets flanked by incongruent distractors (e.g., >><>>) than for targets flanked by congruent distractors (e.g., >>>>>). The SCE refers to the fact that a congruency effect on the current trial is typically larger if the previous trial was congruent than if the previous trial was incongruent (for hypothetical data from the flanker task representing this pattern, see the left panel of Figure 1). Grundy et al. (2017) found no difference between bilinguals and monolinguals in the magnitude of the conventional congruency effect, but a smaller SCE for bilinguals relative to monolinguals (i.e., for bilinguals, there was no significant difference between the congruency effect measured after a congruent versus an incongruent trial; but for a review of several failures to replicate this pattern, see Paap et al., 2019).

According to Grundy et al. (2017), these findings reflect differences in attentional disengagement, an aspect of executive attention that terminates attention to preceding stimuli and diminishes their impact on current processing. Owing to the challenges of managing two languages, bilinguals have considerable experience disengaging attention from preceding stimuli and, hence, are relatively immune to the effect of previous stimuli on current processing. Monolinguals, by contrast, are not. The impact of bilingual experience, however, is confined to attentional disengagement specifically rather than inhibitory processes more broadly and is therefore evident in sequential but not conventional congruency effects. According to Grundy et al., these findings converge with evidence linking individual differences in conflict adaptation with individual differences in executive-attention processes (e.g., van Steenbergen et al., 2015). Further, they explain the widespread null effects of language status on conventional interference measures by suggesting that, being a reflection of inhibitory processes, conventional interference measures are insensitive to the effects of language status.

Although potentially compelling, there are a few weaknesses in Grundy et al.'s (2017) argument. First, attentional disengagement has not been previously validated or even hypothesised as a basis for explaining individual differences in conflict adaptation effects such as the SCE (for review, see Egner, 2014). Thus, any claims that differences in the SCEs of monolinguals and bilinguals reflect differences in attentional disengagement are largely suppositional (see also Goldsmith & Morton, 2018a, 2018b). In fact, some of the evidence from other paradigms that Grundy et al. (2017) cited in support of a difference in attentional disengagement between monolinguals and bilinguals (Mishra et al., 2012) have also failed to replicate in contexts that are different from the original one (Saint-Aubin et al., 2018). Second, Grundy et al. (2017) treated language status as a dichotomous variable, with monolinguals categorized into one group and bilinguals into another. As discussed earlier, treating language status categorically can, on some accounts, obscure subtle relations between multilingualism and higher-order cognition.

The goal of the present research was therefore to undertake a more principled examination of bilingualism and conflict adaptation effects. To achieve that goal, we first grounded our investigation in validated theoretical terms by examining language status effects through the lens of the Dual Mechanisms of Control (DMC) framework (Braver, 2012; Braver et al., 2007), a computationally and neurophysiologically-validated model of conflict adaptation and related individual differences. According to the DMC framework, individuals rely on two modes of control when managing conflict: a proactive mode of control that involves the prospective anticipation of forthcoming conflict via an incremental strengthening of actively maintained attention-guiding rules; and a reactive mode of control that is supported by episodic memory and is set in motion "on the fly" in response to unanticipated events. Second, we treated language status as a continuous rather than as a dichotomous variable and measured it alongside continuous variation in WM capacity, an established correlate of individual differences in conflict adaptation effects and executive attention (e.g., Kane & Engle, 2003). Doing so allowed us to replicate known conflict adaptation effects and patterns of individual differences therein, and test whether those patterns extended to language status.

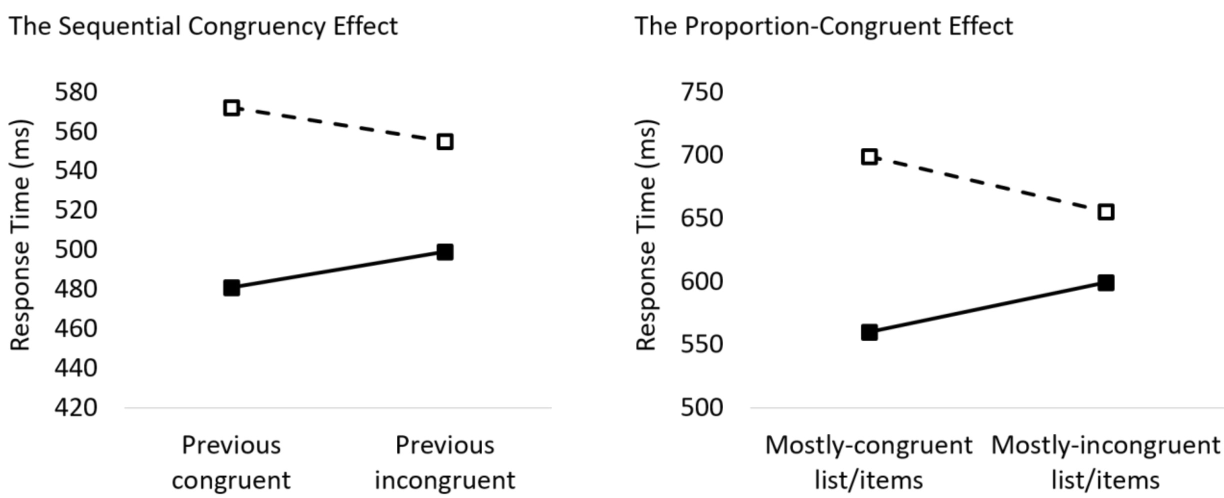
We conducted our examination of these issues in two separate experiments. In Experiment 1, individual differences in proactive control were examined through the use of a list-wide proportion-congruent effect (list-wide PCE). In Experiment 2, individual differences in reactive control were examined through the use of an item-specific PCE. In tasks such as the Stroop (1935) task, the list-wide PCE (Logan & Zbrodoff, 1979) refers to differences in congruency effects that occur with changes in the proportion of congruent items in a list of trials (for hypothetical data from the Stroop task representing this pattern, see Figure 1, right panel; note that the figure also represents the fact that latencies are typically longer in the Stroop task compared to the flanker task). In lists with more incongruent than congruent items (Mostly-Incongruent [MI] lists), congruency effects are typically smaller than in lists with more congruent than incongruent items (Mostly-Congruent [MC] lists). This pattern is typically presumed to reflect a form of conflict adaptation involving a proactive mode of control. List-level saturation favouring incongruent trials supports the sustained maintenance of attention-guiding rules, leading to advanced preparation for and attenuated costs of conflict. The item-specific PCE (Jacoby et al., 2003), on the other hand, refers to changes in the magnitude of the congruency effect across specific pairs of stimuli in a list. For one pair (e.g., the words “RED” and “BLUE” and the colors red and blue), the incongruent stimuli are set up to occur more frequently than the congruent, whereas for the second pair (e.g., the words “GREEN” and “YELLOW” and the colors green and yellow), the opposite is true. When interleaved, pairs of this kind lead to a list-wide congruency proportion of 50%, meaning that it is impossible to anticipate from trial to trial whether a forthcoming stimulus will be congruent or incongruent. Nevertheless, congruency effects are smaller for stimulus pairs for which the incongruent stimulus is more frequent

(Mostly-Incongruent [MI] items) than for pairs for which the congruent stimulus is more frequent (Mostly-Congruent [MC] items). Adaptation effects of this kind are typically assumed to reflect a reactive mode of control because there is no basis for anticipating the need for conflict processing in lists of this kind. By measuring list-wide and item-specific PCEs, Experiments 1 and 2 tested the association between language status and proactive and reactive modes of control, respectively.

To ensure that we could detect and interpret any subtle associations between language status and these forms of conflict adaptation, language status was measured continuously as was variation in WM capacity. Because language status – like WM capacity – has been linked to variation in executive attention, the expectation would be that language status and WM capacity should relate comparably to measures of conflict adaptation.

Figure 1

The typical Sequential Congruency Effect and Proportion-Congruent Effect



Note. The solid line and the dotted line represent latencies for congruent and incongruent items, respectively. Represented in the left and right panels are typical but hypothetical patterns of data observed in conventional interference paradigms like the Stroop task or the Eriksen flanker task, respectively.

Experiment 1

The goal of Experiment 1 was to test the association between bilingualism and proactive control, a form of conflict adaptation that involves the sustained maintenance of task-relevant information (Braver, 2012; Braver et al., 2007). Participants of varying second-language proficiency (measured on a continuous dimension) were administered a list-wide PCE paradigm. Color-word Stroop stimuli were presented in two different lists that differed in conflict saturation: an MC list with 25% incongruent trials and an MI list with 75% incongruent trials. Language status was measured by means of a comprehensive language-use questionnaire (Paap & Greenberg, 2013). WM capacity was assessed by means of a battery of complex span tasks, as a battery provides a more reliable estimate of WM capacity than does a single task (Foster et al., 2015).

The color-word Stroop implementation of the list-wide PCE paradigm has been widely used to examine individual differences in executive attention, particularly in relation to WM capacity. Of special interest, Kane and Engle (2003; see also Hutchison, 2011) found reduced Stroop interference in high compared to low WM-capacity individuals. This pattern emerged more clearly in MC lists and was interpreted as reflecting the fact that high WM-capacity individuals, but not low WM-capacity individuals, can actively maintain attention-guiding rules even in contexts which do not support their active maintenance, such as MC lists. This pattern was obtained using an extreme-groups comparison but was later replicated by Meier and Kane (2013) using a continuous measure of WM capacity, with the exception that there was little indication in that case that WM-related differences were more pronounced in MC lists than in MI lists. In general, though, available evidence suggests that individuals with higher WM capacity exhibit smaller congruency effects relative to individuals with lower WM capacity, with potentially more pronounced differences in MC lists.

We used these well-replicated effects as a baseline in an examination of the relation between language status and proactive control. If differences in bilingualism (as expressed by such measures as second-language proficiency) predict variance in executive attention much as differences in WM capacity do, then increasing bilingualism should be associated with decreasing congruency effects, especially in MC lists.

Method

Participants

One hundred and fifty-seven undergraduate students were recruited through the Western University research participant pool (see SONA recruitment posting in the Supplementary Materials). Of these, we removed 3 participants because of an excessive number of errors and null responses on the Stroop task (above 25%), 3 participants because they erroneously were given more than one block of one or more of the complex span tasks (see below), 2 participants because they failed to complete one or more of those tasks, one participant who failed to complete the questionnaire, and one participant whose Stroop task data were not recorded. Further, in line with traditional practice in individual-differences research (e.g., Unsworth et al., 2005), 22 participants were also removed because their accuracy on the distractor component of one or more of the complex span tasks was below 75%. For example, a participant was excluded if, in the Operation Span task (one of the complex span tasks that we used), they responded correctly to less than 75% of the math questions (the distractor component of that task) even if they simultaneously recalled all letters (in the correct order) that were presented between the math questions (the memory component of that task from which WM-capacity scores are derived).

An exclusion criterion of this sort is often enforced to guard against the possibility of including in the analyses participants who strategically decreased their attention to the distractor component of a complex span task (e.g., those who put little effort into responding accurately to the math questions) in order to artificially improve their performance on the memory component (e.g., to recall more letters in the correct order). Memory scores of these participants are likely an overestimate of their actual WM capacity, a construct which, by definition, is the ability to actively maintain task-relevant information *in the face of distraction* (Kane & Engle, 2003). The exclusion criterion is thus aimed to ensure that most of the participants included in the analyses did indeed face distraction (i.e., engaged in the distractor component of the task as they were supposed to). Note that, for this and the following experiment, we used a 75% cut-off for performance on the distractor component of any of the complex span tasks because we found it to be an acceptable compromise between the 85% cut-off commonly used in early research with complex span tasks (e.g., Unsworth et al., 2005) (which resulted in the exclusion of a large number of participants, i.e., 49 of our initial 157 participants, thus severely limiting the statistical power of the analysis), and more recent trends for using more liberal or even no cut-offs based on evidence that doing so does not change the psychometric properties of complex span tasks in any real way (e.g., Đokić et al., 2018; Unsworth et al., 2009; for a similar cut-off, see Spinelli et al., 2020).

The final sample included 125 total participants. Of this sample, 85 participants were females and 40 were males, and 105 rated their English speaking proficiency as better than or equal to their proficiency in any other language, with the remaining 20 participants still rating their English speaking proficiency as high (*M* = 5.15, *SD* = .75, range 4-6, on a scale where 5 indicated "near fluency" and 6 indicated a level of fluency comparable to that of a native speaker; removing these participants did not change the pattern of results reported below). The typical education level was graduation from high school (118), with the remaining 7 participants having higher education levels, up to a Bachelor's Degree. Additional descriptive information about the final sample is reported in Table 1. A breakdown of the second languages listed, i.e., the second most proficient language listed by each participant based on speaking proficiency, is reported in Table S1. All participants had normal or corrected-to-normal vision and received course credit for their participation.

Table 1

Characteristics of participants in Experiments 1 and 2

	Experiment 1			Experiment 2		
Characteristic	Mean	SD	Range	Mean	SD	Range
Age	18.82	1.43	17-27	18.93	1.38	18-27
Years in Canada	14.22	6.80	0-27	17.63	3.53	5-27
Number of languages reported	2.47	.96	1-5	1.89	.95	1-5
First-language speaking proficiency	6.68	.49	5-7	6.56	.50	6-7
Second-language speaking proficiency	3.83	2.07	0-7	1.97	2.10	0-6
First-language age of exposure	0.91	2.01	0-9	.42	1.31	0-7
Second-language age of exposure	3.07	4.19	0-19	4.59	4.32	0-17
First-language monthly use	79.62	21.94	6-100	93.35	14.90	5-100
Second-language monthly use	17.27	19.31	0-80	5.93	14.31	0-95
Language switch frequency	2.61	1.15	1-5	1.77	1.01	1-5
Family income	3.54	.85	1-5	3.76	.86	1-5
WM capacity	9.36	2.08	3.23-13.63	10.08	1.75	5.01-13.67

Note. “First language” and “second language” refer to the first and second most proficient language listed, respectively. Speaking proficiency was measured on a 7-point Likert scale question, whereas switch frequency (based on 3 questions about language switching behavior) and family income were measured on 5-point Likert scales. Age of exposure refers to the age (in years) at which participants were first exposed to the language in question. Monthly use was expressed in terms of percentage of time (8 participants for which the percentages did not add up to 100 were excluded from this calculation). For the second-language items, a score of 0 was assigned to participants who listed only one language (except for the language switch frequency item, where a score of 1 indicates no switching). Second-language speaking proficiency was the score used for the Bilingualism variable. See the main text for an explanation of the WM-capacity score.

Materials

For the Stroop task, 4 color words (RED, BLUE, GREEN, YELLOW) were used as carrier words and the corresponding colors (red [R: 255; G: 0; B: 0], blue [R: 0; G: 112; B: 192], green [R: 0; G: 176; B: 80], and yellow [R: 255; G: 255; B: 0], corresponding to “red”, “blue”, “green” and “yellow” in the standard DMDX palette; Forster & Forster, 2003) were used as targets. Each block included 96 trials. In the MC block, each word appeared in the congruent color 75% of the time and in an incongruent color 25% of the time. Conversely in the MI block, each word appeared in the congruent color 25% of the time and in an incongruent color 75% of the time. Note that each word was not combined with each of the four colors. Instead, the words RED and BLUE were presented only in red and blue ink colors, and the words GREEN and YELLOW were presented only in green and yellow ink colors.

The reason we used this particular design was to allow for a process that participants could potentially engage during the Stroop task, contingency learning (Schmidt, 2013, 2019; Schmidt & Besner, 2008). In color identification tasks, contingency learning refers to the process of learning to associate individual words with their typical response (Schmidt et al., 2007; Lin & MacLeod, 2018). In classic list-wide PC manipulations in the Stroop task, this process is always possible in MC lists (e.g., participants might learn to associate the word RED with the response red, since RED occurs in the color red most of the time). However, contingency learning is not always possible in MI lists, particularly when more than two colors are used and each word is presented in each of those colors (e.g., in an MI list, participants would not learn to associate the word RED with any response if RED appears in red, blue, green, and yellow equally often). Although there is increasing evidence that the role that contingency learning plays in both list-wide and item-specific PC manipulations is relatively minor (e.g., Spinelli & Lupker, 2020a, 2020b; Spinelli et al., 2020; Spinelli et al., 2019), it is not negligible either. Thus, we opted for a “symmetric” design, i.e., one where contingency learning was possible in the MI block (e.g., participants could learn a RED-blue association) to the same extent that it was possible in the MC block (e.g., participants could learn a RED-red association). With this design, if contingency learning were to impact performance, it would do so similarly in both the MC and MI blocks.

To measure WM capacity, we used a battery of complex span tasks including the Operation Span task, the Symmetry Span task, and the Rotation Span task (Conway et al., 2005; Kane et al., 2004; Redick et al., 2012; Unsworth et al., 2005). These tasks were shortened versions of complex span tasks (one block instead of the regular three blocks) aimed to test different constructs in working memory, so as to obtain reliable measures of WM capacity as a whole while minimizing testing duration (Foster et al., 2015).

Finally, we used the background questionnaire developed by Paap and collaborators (e.g., Paap & Greenberg, 2013) to obtain information about participants' demographics, language status, socio-economic status, and other variables known to influence executive functioning.

Procedure

Participants performed the study in a single, 1-hour session. They first completed the Stroop task. In this task, each trial began with a 250-ms fixation symbol ("+") followed by a 250-ms blank screen. In the next display, a colored word appeared in uppercase for 2000 ms or until the participant's response. Participants were instructed to ignore the word and identify the color as quickly and as accurately as possible and to respond by pressing the "J" key for red, the "K" key for blue, the "L" key for green, and the ";" key for yellow using the four fingers of their right hand. Finally, there was a 300-ms blank screen followed by a 500-ms feedback message, in white font, which read "Correct", "Incorrect" or "No response" for correct, incorrect, or missed responses, respectively. All stimuli were presented in Courier New font, pt. 14, against a medium grey background (R: 169; G: 169; B: 169). There was a self-paced pause between the first and the second block. The order of trials within each block was randomized and the order of presentation of the two blocks was counterbalanced across participants. Prior to starting the first block, participants performed a practice session of 16 trials, of which 8 trials were congruent and 8 incongruent. They were told nothing about the congruency proportion of either the first or the second block. This task was run using DMDX (Forster & Forster, 2003) software.

Next, participants completed the battery of shortened complex span tasks. They completed one block of the Operation Span task, followed by one block of the Symmetry Span task, followed by one block of the Rotation Span task. In these tasks, participants were given a sequence of to-be-remembered items (e.g., a sequence of letters) and had to complete a distractor task (e.g., solving a math problem) between the presentations of each of the to-be-remembered items in the sequence. The sequence of to-be-remembered items varied from two to five items (Symmetry Span and Rotation Span tasks) or from three to seven items (Operation Span task). Scores are calculated by summing the number of items correctly recalled in the correct order, a measure known as the partial score (Turner & Engle, 1989; for more details, see Conway et al., 2005; Kane et al., 2004; Redick et al., 2012; Unsworth et al., 2005). These tasks were run using E-prime (Psychology Software Tools, Pittsburgh, PA) software.

Finally, participants completed the background questionnaire, which was implemented with Qualtrics (Qualtrics, Provo, UT) software and administered through an internet browser. For this experiment and for Experiment 2, all participants gave informed consent before taking part in the study. The research was approved by the Research Ethics Board of the University of Western Ontario (protocols #112910 and #108956 for Experiments 1 and 2, respectively).

Results

Prior to the analyses, responses faster than 300 ms or slower than the time limit (i.e., 2000 ms) on the color Stroop task (accounting for 0.9% of the data) were discarded. Latency analyses were conducted only on correct responses.

The analysis was conducted using mixed-effects modelling, a type of analysis which permits use of both continuous and categorical variables (the fixed effects), while controlling for variance among the participants and the items being used (the random effects; Baayen, 2008; Baayen et al., 2008; for similar analyses in the context of the Stroop task, see Meier & Kane, 2013, 2015; Spinelli et al., 2020). Latencies and errors were analyzed using generalized linear mixed-effects models (GLMMs) in R version 3.5.1 (R Core Team, 2018), treating subjects, colors, and words as random effects. The fixed effects were Congruency (congruent vs. incongruent), Block Type (MC vs. MI), Order (MC first vs. MI first), plus two continuous measures for WM Capacity and Bilingualism, respectively. The measure for WM Capacity was obtained by scaling the partial score for the Operation Span Task (which can range from 0 to 25 in the one-block version of that task) to fit the scale of the Symmetry and Rotation Span Tasks (which can range from 0 to 14 in the one-block version of those tasks). Consistent with the idea that individual complex span tasks measure task-specific abilities (e.g., ability at solving math problems) in addition to WM Capacity (e.g., Loehlin, 2004), the scores for the three complex span tasks were only weakly correlated with one another (Operation-Symmetry, $r = .129$, $p = .125$, Operation-Rotation, $r = .336$, $p < .001$, Symmetry-Rotation, $r = .317$, $p < .001$), Cronbach's $\alpha = .491$. The three scores were therefore averaged to obtain a single composite score for each participant (Conway et al., 2005). The measure for Bilingualism was the self-rated speaking proficiency for the second most proficient language, with a score of 0 assigned to participants who listed only one language. Excluding participants who listed only one language, speaking, understanding, writing, and reading skills for the second most proficient language were moderately to strongly correlated with one another, range: .451-.882, all *p*s < .001, Cronbach's $\alpha = .891$. Speaking proficiency was used as the measure for Bilingualism because (second) language production appears to be the primary domain in which inhibitory control is engaged (Green, 1998). Relative to the other language skills, speaking proficiency also generally correlated more strongly (in the expected direction) with other important measures of bilingualism such as age of exposure to the second language (r = -.219, p = .023), monthly use of the second language (r = .619, p < .001), and language switch frequency (r = .544, p < .001). Notably, however, this measure of Bilingualism did not correlate with the measure used for WM Capacity (i.e., the composite score of the three complex span tasks), r = -.09, p = .295.

Prior to running the model, R-default treatment contrasts were changed to sum-to-zero contrasts (i.e., contr.sum) to help interpret lower-order effects in the presence of higher-order interactions (Levy, 2014; Singmann & Kellen, 2018). The lme4 package, version 1.1-18-1 (Bates et al., 2015) was used to run the GLMMs. The models were fit by maximum likelihood with the Laplace approximation technique. Model estimation was conducted using the BOBYQA optimizer, an optimizer known to generate fewer false-positive convergence failures than other optimizers in the current version of lme4, with a maximum number of 1,000,000 iterations. Continuous predictors (i.e., WM Capacity and Bilingualism) were also centered and scaled prior to the analyses to help model estimation (Bolker, 2020). The emmeans package, version 1.3.1 (Lenth, 2018), was used to conduct follow-up tests. The ggplot2 package, version 3.1.0 (Wickham, 2016), was used to generate graphs. A Gamma distribution was used to fit the raw RTs, with an identity link between the fixed effects and the dependent variable (Lo & Andrews, 2015; Yang et al., 2019), whereas a binomial distribution with a logit link between the fixed effects and the dependent variable was used to fit the error data. The mean RTs and error rates based on subject means are presented in Table 2. For this and the following experiment, the raw data and R and SPSS files used for the analyses are publicly available at https://osf.io/gz6h3/. The study materials are available upon request. Neither study was preregistered.

Table 2

Mean RTs and error rates (and corresponding standard errors) for Experiment 1

	RTs		Error rates	
Congruency	MC block	MI block	MC block	MI block
MC first				
Congruent	712 (16)	710 (17)	.019 (.004)	.028 (.005)
Incongruent	890 (23)	775 (17)	.081 (.011)	.044 (.005)
Congruency Effect	178	65	.062	.016
MI first				
Congruent	691 (12)	728 (13)	.029 (.004)	.031 (.006)
Incongruent	819 (16)	790 (16)	.078 (.008)	.048 (.005)
Congruency Effect	128	62	.049	.017

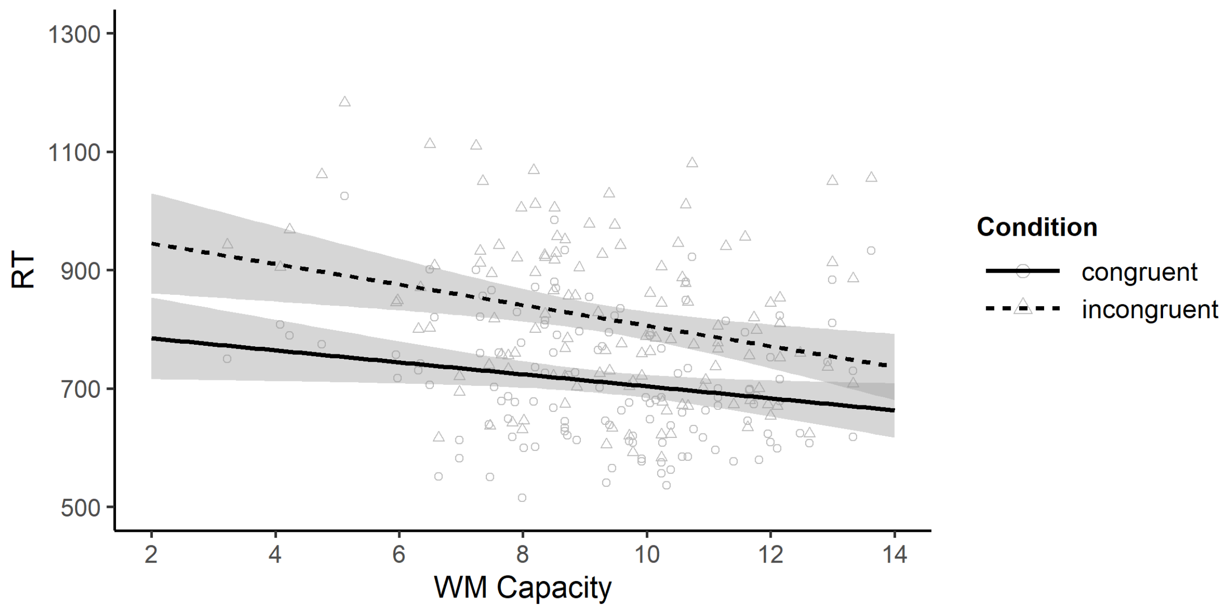
RTs

The entire set of results from the latency analyses are reported in Table S2 in the Supplementary Materials. Here, we focus on the most relevant results. As expected, the overall effect of Congruency, $\beta = -51.40$, $SE = 1.30$, $z = -39.54$, $p < .001$, was subject to a list-wide PCE, such that congruency effects were smaller for MI compared to MC blocks, $\beta = -21.43$, $SE = 1.50$, $z = -14.28$, $p < .001$. This list-wide PCE was further modulated by the order in which the blocks were presented, such that the list-wide PCE was larger in the MC-first than the MI-first condition. This pattern was confirmed by 3-way interaction between Congruency, Block Type, and Order, $\beta = -4.92$, $SE = 1.29$, $z = -3.82$, $p < .001$. Follow-up tests indicated that MC blocks elicited smaller congruency effects in the MI-first condition (128 ms) than in the MC-first condition (178 ms), $\beta = -52.38$, $SE = 7.41$, $z = -7.07$, $p < .001$, whereas congruency effects in MI blocks were only marginally smaller in the MI-first condition (62 ms) than in the MC-first condition (65 ms), $\beta = -13.01$, $SE = 7.04$, $z = -1.85$, $p = .065$. This pattern, known as the asymmetric list-shifting effect, replicates the original pattern reported by Abrahamse et al. (2013).

Of primary interest in the analysis were individual differences in the magnitude of congruency effects generally and list-wide PCEs more specifically. With regards to WM Capacity, congruency effects decreased with increasing WM Capacity (see Figure 2), as confirmed by a 2-way interaction between Congruency and WM Capacity, $\beta = 7.64$, $SE = 1.40$, $z = 5.46$, $p < .001$. There was, however, no evidence of an association between the list-wide PCE and WM Capacity, as can be seen in Figure 3. Indeed, the 3-way interaction of Congruency, Block Type, and WM Capacity, $\beta = -.54$, $SE = 1.36$, $z = -.40$, $p = .692$, was not significant.

Figure 2

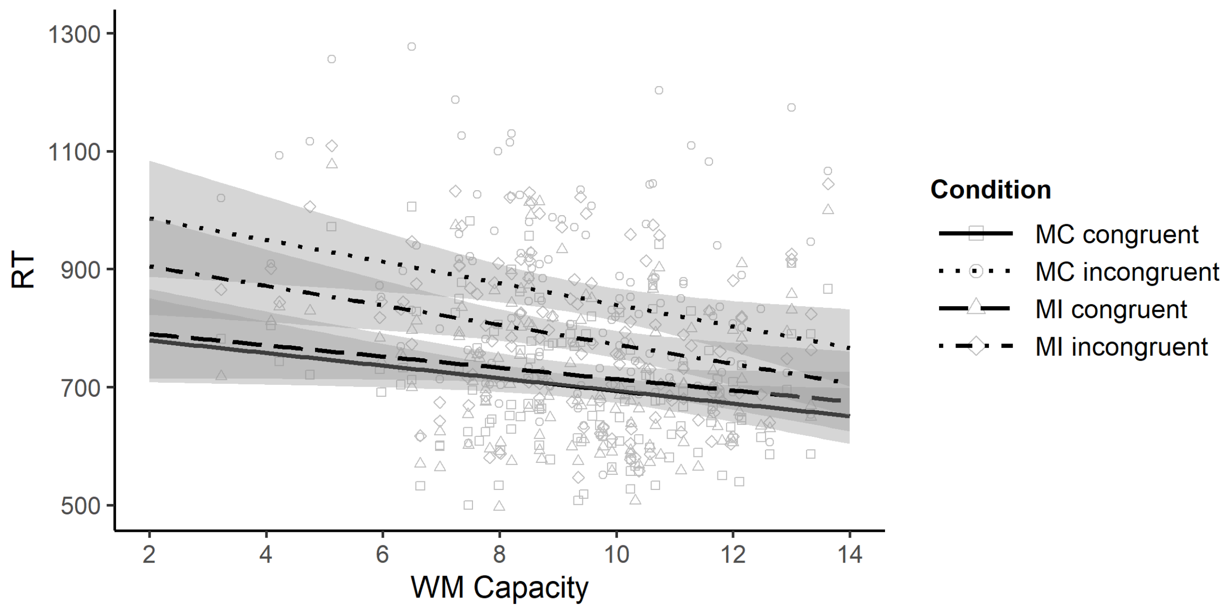
The impact of WM capacity on the congruency effect in latencies for Experiment 1



Note. A scatterplot of participants' WM capacity and their mean latencies to congruent (circles) and incongruent (triangles) items. Regression slopes (with 95% confidence interval bands) for congruent and incongruent items are marked with a solid line and a dashed line, respectively. When moving from the left (lower WM capacity) to the right side of the graph (higher WM capacity), there is a decrease both in overall response latency and the congruency effect (i.e., the distance between the solid line [congruent] and the dashed line [incongruent items]).

Figure 3

The impact of WM Capacity on the list-wide PCE in latencies for Experiment 1

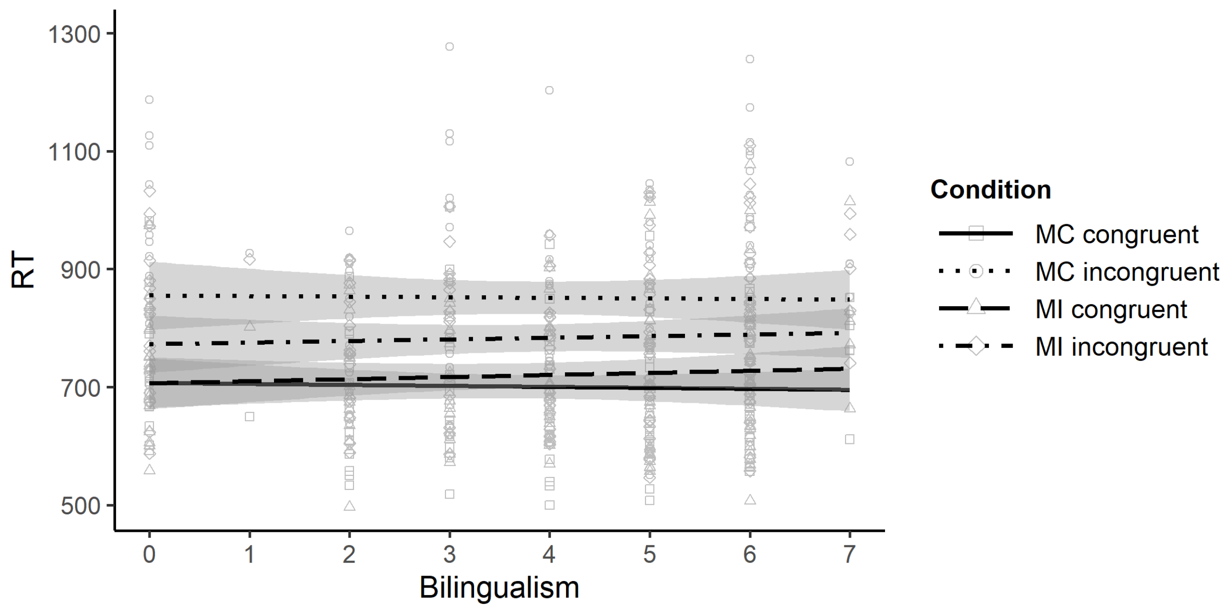


Note. A scatterplot of participants' WM capacity and their mean latencies to congruent items in MC blocks (squares), incongruent items in MC blocks (circles), congruent items in MI blocks (triangles), and incongruent items in MI blocks (rhombuses). Regression slopes (with 95% confidence interval bands) for congruent items in MC blocks, incongruent items in MC blocks, congruent items in MI blocks, and incongruent items in MI blocks, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

With respect to bilingualism, there was no evidence of associations with response latency, congruency, or the PCE (see Figure 4). As can be seen in Figure 5, however, there was some evidence of a high-level interaction between bilingualism and the PCE in the form of a 4-way interaction between Congruency, Block type, Order, and Bilingualism, $\beta = -2.60$, $SE = 1.29$, $z = -2.02$, $p = .043$. Specifically, increasing Bilingualism was associated with decreasing congruency effects in the MI block in the MC-first condition, $\beta = 15.64$, $SE = 5.57$, $z = 2.81$, $p = .005$, but not in other situations. This pattern was driven by the fact that in the MI block of the MC-first condition, increasing Bilingualism was associated with increasing latencies to congruent items, $\beta = 14.20$, $SE = 4.43$, $z = 3.20$, $p = .001$, but had no impact on latencies to incongruent items, $\beta = -1.44$, $SE = 4.06$, $z = -.36$, $p = .722$. This pattern can be appreciated in the left panel in Figure 5, where the dashed line, the line corresponding to congruent items in the MI block in the MC-first condition, tends to go up, whereas for all other conditions and trial types, the best fitting lines were essentially flat.

Figure 4

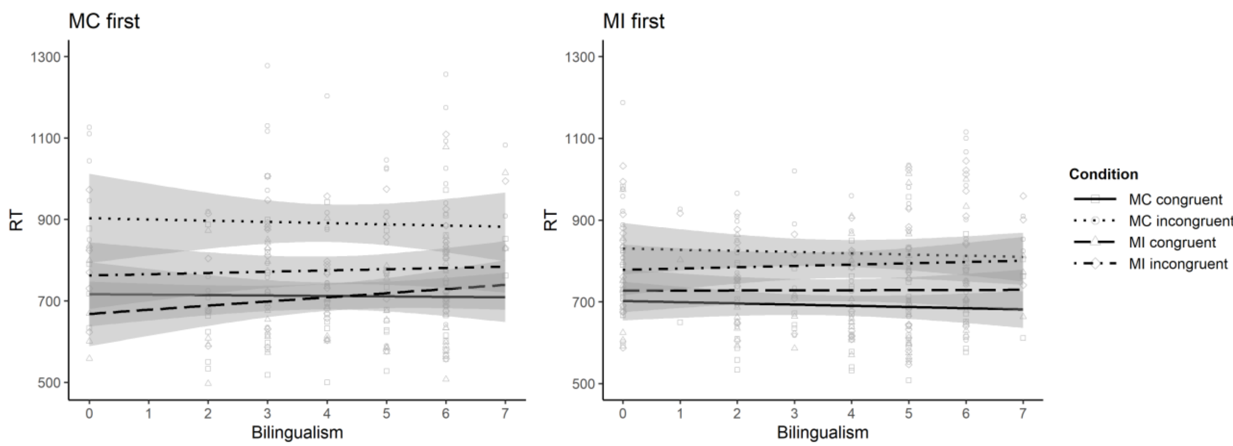
The impact of Bilingualism on the list-wide PCE in latencies for Experiment 1



Note. A scatterplot of participants' bilingualism scores and their mean latencies to congruent items in MC blocks (squares), incongruent items in MC blocks (circles), congruent items in MI blocks (triangles), and incongruent items in MI blocks (rhombuses). Regression slopes (with 95% confidence interval bands) for congruent items in MC blocks, incongruent items in MC blocks, congruent items in MI blocks, and incongruent items in MI blocks, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Figure 5

The impact of Bilingualism on the congruency effect in latencies for Experiment 1, examined separately for the MC-first and MI-first conditions



Note. A scatterplot of participants' bilingualism scores and their mean latencies to congruent items in MC blocks (squares), incongruent items in MC blocks (circles), congruent items in MI blocks (triangles), and incongruent items in MI blocks (rhombuses), in the MC-first (left panel) and MI-first (right panel) conditions. Regression slopes (with 95% confidence interval bands) for congruent items in MC blocks, incongruent items in MC blocks, congruent items in MI blocks, and incongruent items in MI blocks, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Errors

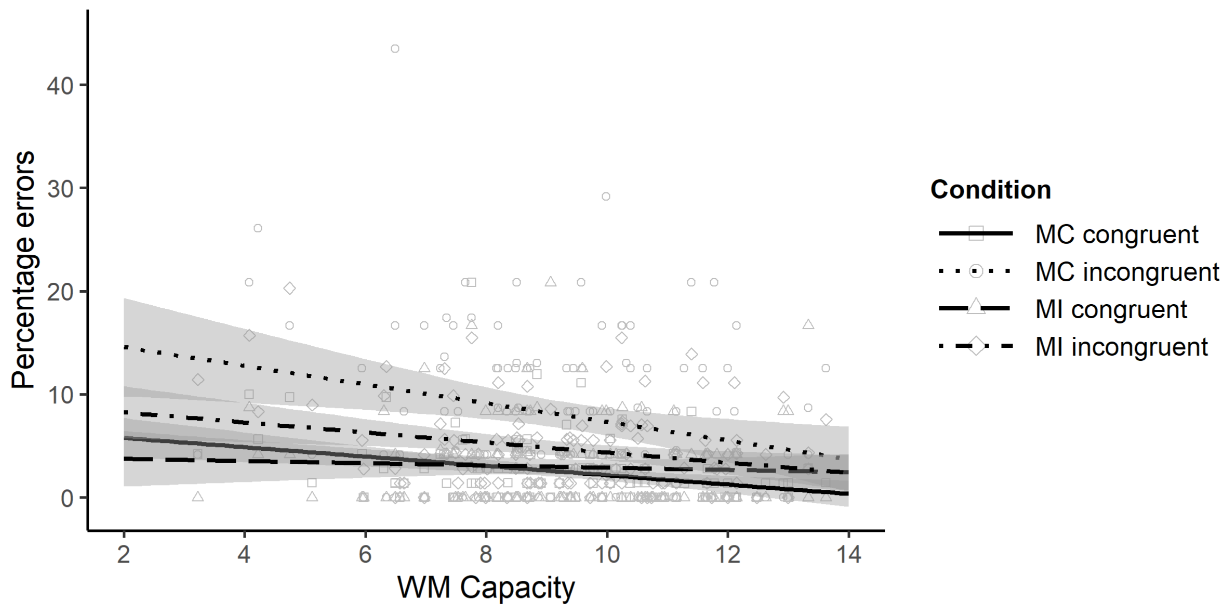
The results of the error analyses are reported in Table S3 in the Supplementary Materials. Again, here we focus on the most relevant results. There were main effects of Congruency (congruent more accurate than incongruent), β = .49, *SE* = .04, z = 11.61, p < .001, Block Type (MI more accurate than MC), β = -.09, *SE* = .04, z = -2.16, p = .031, and, as expected, an interaction between Congruency and Block Type, β = .21, *SE* = .04, z = 4.97, p < .001, indicating a regular list-wide PCE: Congruency effects were larger in MC blocks than in MI blocks. There was also a small asymmetric list-shifting effect as reflected by a marginal interaction between Congruency, Block Type, and Order, β = .07, *SE* = .04, z = 1.74, p = .082. As expected (Abrahamse et al., 2013), list-wide PCEs (defined as the difference between congruency effects in MC blocks and congruency effects in MI blocks) were slightly larger in the MC-first condition (MC block = 6.2%; MI block = 1.6%; PCE = 4.6%) than in the MI-first condition (MC block = 4.9%; MI block = 1.7%; PCE = 3.2%), with this difference mainly deriving from MC blocks.

Of primary interest in the analysis were individual differences in list-wide PCE effects. With regards to WM capacity and the list-wide PCE, we did find a three-way interaction between Block Type, Congruency, and WM Capacity, β = .08, *SE* = .04, z = 2.04, p = .041. Follow-up tests revealed that the source of this interaction was the fact that WM Capacity had no impact on congruent items in MI blocks, β = .04, *SE* = .13, z = .30, p = .768, whereas higher WM Capacity was associated with higher accuracy in all other conditions (i.e., congruent and incongruent items in MC blocks and incongruent items in MI blocks). This pattern is represented in Figure 6, with the regression line for congruent items in the MI blocks (the dashed line) being the only one that does not go down with higher WM Capacity. WM Capacity also interacted with Order, β = .17, *SE* = .07, z = 2.33, p =.020. Follow-up tests revealed that higher WM Capacity was associated with higher accuracy in the MC-first condition, β = .40, *SE* = .11, z = 3.64, p < .001, but not significantly so in the MI-first condition, β = .07, *SE* = .09, z = .80, p = .423.

With regards to Bilingualism and the list-wide PCE, the 3-way interaction between Congruency, Block Type, and Bilingualism was not significant, $\beta = -.02$, $SE = .04$, $z = -.36$, $p = .717$, indicating that Bilingualism did not alter the list-wide PCE. This pattern is represented in Figure 7. Bilingualism was, however, involved in a 3-way interaction with Congruency and Order, $\beta = .09$, $SE = .04$, $z = 2.02$, $p = .043$. Follow-up tests revealed that Bilingualism had opposite effects on accuracy for congruent items in MC-first vs. MI-first conditions, $\beta = .53$, $SE = .19$, $z = 2.83$, $p = .005$, as higher Bilingualism led to marginally higher accuracy on congruent items in the MC-first condition, $\beta = .28$, $SE = .15$, $z = 1.96$, $p = .050$, whereas it led to lower accuracy on congruent items in the MI-first condition, $\beta = -.25$, $SE = .12$, $z = -2.07$, $p = .038$. Bilingualism was unrelated to accuracy on incongruent items in either the MC-first condition, $\beta = .05$, $SE = .12$, $z = .43$, $p = .671$, or the MI-first condition, $\beta = -.14$, $SE = .09$, $z = -1.46$, $p = .144$.

Figure 6

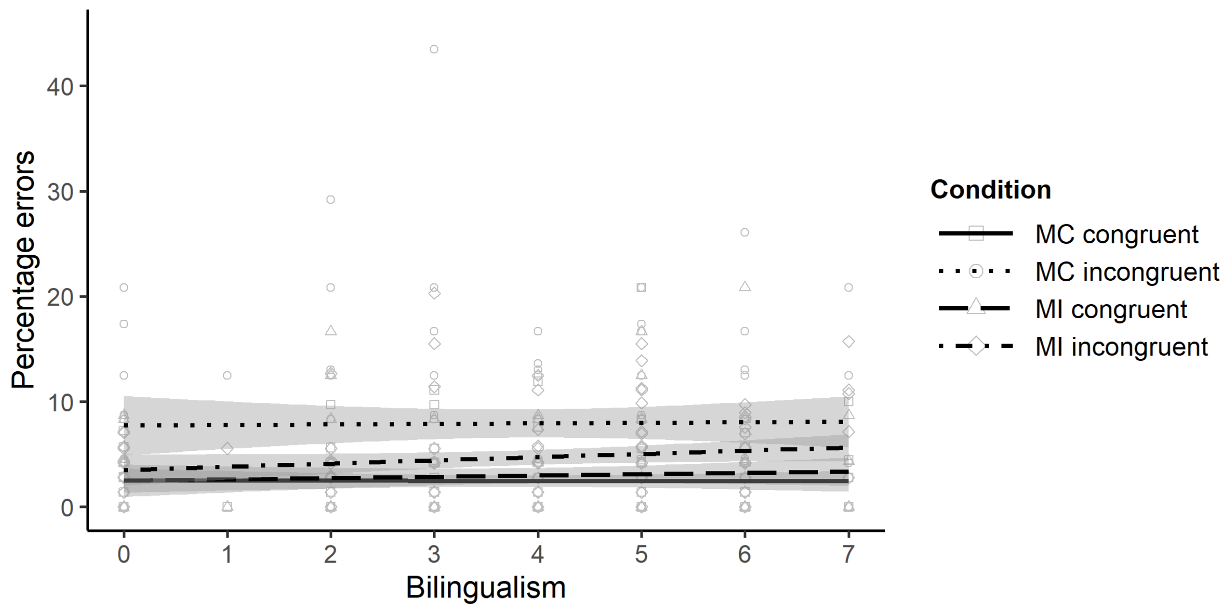
The impact of WM Capacity on the list-wide PCE in error rates for Experiment 1



Note. A scatterplot of participants' WM capacity and their mean error rates for congruent items in MC blocks (squares), incongruent items in MC blocks (circles), congruent items in MI blocks (triangles), and incongruent items in MI blocks (rhombuses). Regression slopes (with 95% confidence interval bands) for congruent items in MC blocks, incongruent items in MC blocks, congruent items in MI blocks, and incongruent items in MI blocks, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Figure 7

The impact of Bilingualism on the list-wide PCE in error rates for Experiment 1



Note. A scatterplot of participants' bilingualism scores and their mean error rates for congruent items in MC blocks (squares), incongruent items in MC blocks (circles), congruent items in MI blocks (triangles), and incongruent items in MI blocks (rhombuses). Regression slopes (with 95% confidence interval bands) for congruent items in MC blocks, incongruent items in MC blocks, congruent items in MI blocks, and incongruent items in MI blocks, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Finally, there was an interaction between Congruency, Block Type, Order, WM Capacity, and Bilingualism, β = .10, *SE* = .04, *z* = 2.75, *p* = .006. To explore this five-way interaction, we conducted separate analyses for every combination of Block Type and Order (MC block in MC first, MI block in MC first, MC block in MI first, and MI block in MI first). The MI block in the MI-first condition was the only case where an interaction between Congruency, WM Capacity, and Bilingualism emerged, β = .24, *SE* = .08, *z* = 2.93, *p* = .003. Follow-up tests evaluating the impact of WM Capacity on congruent vs. incongruent trials for participants scoring low vs. high in Bilingualism (as determined by a median split) in that block revealed that the source of this interaction was that higher WM Capacity tended to reduce accuracy on congruent items at low levels of Bilingualism, β = -.50, *SE* = .27, *z* = 1.89, *p* = .059. In contrast, higher WM Capacity tended, if anything, to increase accuracy on incongruent items at low levels of Bilingualism, β = .30, *SE* = .15, *z* = 1.98, *p* = .048, and on congruent and incongruent items at high levels of Bilingualism, β = .00, *SE* = .17, *z* = .01, *p* = .989, and β = .03, *SE* = .12, *z* = .26, *p* = .798, respectively.

Discussion

The results of Experiment 1 replicated several findings associated with the list-wide proportion-congruent paradigm. There was a basic overall list-wide PCE, with congruency effects larger in MC than in MI blocks, and a modulation of the PCE by block order, such that PCEs were smaller when the MI block was presented first compared to second (the asymmetric list-shifting effect; Abrahamse et al., 2013). Both effects were expected and are in general agreement with validated cognitive control accounts. On the DMC model for example (Braver, 2012; Braver et al., 2007; see also Kane & Engle, 2003), frequent conflict such as that experienced in an MI list strengthens actively maintained attention-guiding rules, leading to faster responses to incongruent trials compared to an MC list. Further, once established, a proactive mode of control can carry forward and be sustained when conflict becomes less frequent. Consequently, PCEs tend to be smaller when MI lists are presented first rather than second (Abrahamse et al., 2013).

The results were also broadly consistent with previously reported associations between WM capacity and measures of proactive control. Not only was higher WM capacity associated with reduced latencies and error rates overall, but congruency effects decreased with increasing WM capacity, especially in measures of latency. Interestingly, this congruency-effect decrease was equivalent in MC and MI lists. (note 1) Although early reports linked WM capacity to reduced congruency effects mainly (or only) in MC lists (e.g., Hutchison, 2011; Kane & Engle, 2003), those studies were based on extreme groups comparisons (i.e., high versus low WM capacity). More recent studies that used mixed-effects models and continuous measures of WM, like in the present study, reported decreasing congruency effects with increasing WM capacity in both MC and MI lists (Meier & Kane, 2013). Taken together, the findings are consistent with the idea that individual differences in WM capacity are rooted in differences in aspects of executive attention such as active maintenance and interference management. Consequently, individuals with high WM capacity readily adopt a proactive mode of control when faced with frequent conflict (in an MI list) and sustain this mode of control (i.e., avoid goal neglect) also when conflict occurs less frequently (in an MC list).

The most novel and important finding of Experiment 1 is that associations between WM capacity and proactive control did not extend to language status. First, there was no general benefit in speed or accuracy associated with bilingualism. In the latencies, bilingualism was associated with no overall speed-up; in the errors, bilingualism led, if anything, to worse accuracy overall, especially in MI blocks presented first. More importantly, language status was almost completely unrelated to conflict-adaptation measures of proactive control, save for a reduced congruency effect in the MI list of the MC-first condition. Even here though, language status and WM capacity-related reductions in congruency effects differed, with language-status-related reductions driven by slower responses to congruent items (for a similar pattern, see Bialystok et al., 2008, as noted by Hilchey & Klein, 2011) and WM capacity-related reductions driven by faster responses to incongruent items. Only the latter pattern denotes variation in the capacity to manage interference (Paap et al., 2015). (note 2) Worth noting, this absence of any meaningful associations between language status and proactive control was obtained despite our use of a continuous rather than categorical measurement of language status.

This continuous measure was based on second-language speaking proficiency because of the role that inhibitory control appears to have in spoken language production (Green, 1998). However, mechanisms involved in second-language use and language switching have also been proposed as a source of a potential bilingual advantage (e.g., De Bruin et al., 2015; Prior & Gollan, 2011; Verreyt et al., 2016). Thus, we repeated the analyses using second-language monthly use and language switch frequency (see Table 1) as measures of bilingualism. The results were similar to the results of the analyses including second-language speaking proficiency as the measure of bilingualism: First, neither second-language monthly use nor language switch frequency correlated with WM capacity, r = .02, p = .870, and r = -.07 , p = .470, respectively. Further, while WM capacity modulated overall speed, accuracy, and (in the latencies) congruency effects, neither bilingualism measure did so consistently. Language switch frequency was associated with slightly faster latencies, however, it was associated with larger congruency effects in the error rates. Second-language monthly use had no influence on overall performance, the congruency effect, or the list-wide PCE in either the latencies or the error rates.

In sum, Experiment 1 produced evidence of list-wide PCEs and WM capacity-related variability in these effects consistent with evidence already documented in the literature, but it produced little evidence of comparable associations with language status.

Experiment 2

Proactive control is not the only form of control that adapts to changes in conflict frequency. Adaptation to conflict frequency also occurs reactively, with control parameters (focused vs. relaxed attention) set in the moment based on characteristics of the current item rather than set in advance based on characteristics of the current block. In the item-specific PCE paradigm (Jacoby et al., 2003), for example, a pair of MC items and a pair of MI items are intermixed in the same block, so that the overall block-level probability of conflict is fixed at 0.5. Although proactive anticipation of conflict based on block-level statistics is impossible in lists of this kind, congruency effects are typically larger for MC pairs than for MI pairs. These forms of adaptation to conflict frequency are therefore attributed to reactive control, whereby learned recognition of MC items (e.g., the word RED that more frequently appears in the congruent color red than the incongruent color blue) invokes relaxed attention, whereas learned recognition of MI items (e.g., the word GREEN that more frequently appears in the incongruent color yellow than in the congruent color green) invokes focused attention to color. These learned attention settings contribute to differences in interference effects for MC and MI pairs, but cannot be prepared in advance and are, therefore, considered a manifestation of reactive control (Gonthier et al., 2016; Spinelli & Lupker, 2020a, 2020b; Spinelli et al., 2020).

Reactive control, as proactive control, is assumed to be modulated by individual differences in WM capacity (Braver, 2012). Although all individuals presumably have access to reactive control, a bias for engaging this mode of control is expected in individuals with lower WM capacity. In an item-specific PCE paradigm, this bias for reactive control would lead low WM-capacity individuals to rely heavily on the MC/MI nature of the items to select the control setting that should be used at every moment in the task. The result would be a pronounced item-specific PCE in those individuals. High WM-capacity individuals, on the other hand, would not rely so heavily on the MC/MI nature of the items because they tend to proactively maintain focused attention to color information instead. Therefore, high WM-capacity individuals should show a less pronounced item-specific PCE (as a result of reduced reliance on reactive control), as well as a smaller congruency effect in general (as a result of increased reliance on proactive control).

Few studies have addressed the relation between WM capacity and performance in the item-specific PCE paradigm in the Stroop task. Extant evidence is consistent with the pattern just described, albeit this evidence mainly comes from error rates rather than latencies (Hutchison, 2011; see also Spinelli et al., 2020). While low WM-capacity individuals tend to show a marked item-specific PCE in latencies and errors, high WM-capacity individuals typically show such a pattern in latencies but not in errors. That is, in the errors made by high WM-capacity individuals, congruency effects tend be similar for MC and MI items, and smaller than those produced by low WM-capacity individuals.

In Experiment 2, we addressed the question of whether bilingualism would modulate engagement of adaptation to item-specific conflict frequency, a process based on reactive control, in a similar fashion as WM capacity appears to do. To this end, we measured the WM capacity and language status of young adults, as in Experiment 1, after having them engage in an item-specific PCE paradigm. Our rationale was that if bilingualism works in a way that is similar to the way WM capacity works, then it should confer individuals the ability to engage proactive control with ease. Therefore, bilinguals, like high WM-capacity individuals, should not rely heavily on reactive control. As a result, they should show little or no item-specific PCEs compared to monolinguals, individuals who would be more likely to rely on reactive control.

Method

Participants

One hundred and twenty-seven undergraduate students were recruited through the Western University research participant pool (see SONA recruitment posting in the Supplementary Materials). Of these, we removed 1 participant because of an excessive number of errors and null responses on the Stroop task (above 25%), and 3 participants because they failed to complete one or more of the complex span tasks. Twenty-five participants were also removed because their accuracy on the distractor component of one or more of the complex span tasks was below 75%, leaving a final sample of 98 participants. Of the final sample, 64 participants were females and 34 males, and 94 participants rated their English-speaking proficiency better or equal to their proficiency in any other language, with the remaining 4 participants still rating themselves “fluent” or “near fluency” in English (again, removing these participants did not change the pattern of results reported below). The typical education level was graduation from high school (90), with the remaining 8 participants having higher education levels, up to a Bachelor’s Degree. Additional descriptive information about the final sample is reported in Table 1 and a breakdown of the second languages listed is reported in Table S1. All participants had normal or corrected-to-normal vision and received course credit for their participation. Note that these data were collected as part of a larger project investigating the impact of WM load on item-specific PCEs in the Stroop task and contingency-learning effects in a non-conflict task (Spinelli et al., 2020). Specifically, they are the data for the no-load condition in Spinelli et al.’s Experiment 3B.

Materials

For the Stroop task, the same words and colors were used as in Experiment 1. Also similar to Experiment 1, the words RED and BLUE were presented only in red and blue colors, and the words GREEN and YELLOW were presented only in green and yellow colors. In this experiment, however, one set of words (e.g., RED and BLUE) was designated as the MC set and the other set of words (e.g., GREEN and YELLOW) was designated as the MI set. Each word in the MC set appeared in the congruent color 75% of the time and in an incongruent color 25% of the time. Conversely, each word in the MI set appeared in the congruent color 25% of the time and in an incongruent color 75% of the time. The two sets were intermixed in the one 96-trial block so that the trial-to-trial probability of an incongruent trial was fixed at 0.5. The assignment of words to the MC and MI set was counterbalanced across participants.

Either prior to this Stroop block or following it, participants completed another 96-trial block where color-unrelated words, instead of color words, were used (the colors and the color-response mappings were the same as in the Stroop block). The reason for the presence of that non-conflict block was to control for the potential impact of contingency learning in the item-specific PC manipulation in the Stroop task. As noted above when presenting the materials used in Experiment 1, contingency learning refers to the possibility that, in PCE paradigms such as the item-specific PCE paradigm, participants might learn to associate words with their typical response (e.g., a RED-red association), rather than with the control process that is appropriate to the conflict frequency specific to the word (e.g., relaxing attention for RED, an MC word; Schmidt & Besner, 2008). Since it is not obvious whether and how WM capacity or bilingualism would affect this contingency learning process, an examination of performance in that non-conflict block lies beyond the scope of this paper. In addition, as also noted above, the impact of contingency learning in PCE paradigms appears relatively minor. Therefore, we do not discuss the methods or the results for that block here. For this information, the interested reader can consult Spinelli et al. (2020) and Tables S6 (latency analysis) and S7 (error analysis) in the present Supplementary Materials. The rest of the materials (i.e., the battery of complex span tasks and the background questionnaire) were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1, with the exception that there was only one Stroop block. That block was administered either before or after a non-conflict block. The order of the two blocks was counterbalanced across participants. Also, instead of completing 16 practice trials initially, participants completed 8 practice trials prior to the Stroop block. The practice trials mirrored the frequency of word-color combinations in the upcoming block. As in Experiment 1, participants performed the color-identification task (with a Stroop block and a non-conflict block) first, the complex-span task battery second, and completed the background questionnaire at the end.

<u>Results</u>

Prior to the analyses, responses faster than 300 ms or slower than the time limit (i.e., 2000 ms) on the Stroop task (accounting for 0.9% of the data) were discarded. Latency analyses were conducted only on correct responses.

The analysis was conducted in the same way as in Experiment 1, with the exception that the fixed effect of Block Type was replaced with Item Type (MC items vs. MI items) since, in this experiment, the proportion-congruent manipulation was based on a contrast between items intermixed in the same block rather than a contrast between blocks. Note, further, that the fixed effect of Order, in this experiment, referred to whether the Stroop block was presented as the first block (before the non-conflict block) or as the second block (after the non-conflict block). In this experiment as in Experiment 1, it was important to take into account the order in which participants performed the Stroop block to control for effects of practice with the color-response mappings (participants who did that block as the second block had more practice than participants who did that block first, since the colors and the responses made to those colors in the two blocks were the same). As in Experiment 1, the scores for the three complex span tasks (with the scores for the Operation Span task scaled to fit the scale of the other two tasks) were weakly correlated with one another (Operation-Symmetry, $r = .155$, $p = .128$, Operation-Rotation, $r = .140$, $p = .170$, Symmetry-Rotation, $r = .335$, $p = .001$), Cronbach's $\alpha = .445$, and were averaged into a composite score which was used as the measure for WM Capacity. Also as in Experiment 1, excluding participants who only listed one language, speaking, understanding, writing, and reading skills for the second most proficient language were moderately to strongly correlated with one another, range: .630-.879, all ps < .001, Cronbach's $\alpha = .919$, with speaking and understanding proficiency also correlating relatively well with other measures of bilingualism such as age of exposure to the second language ($r = -.225$, $p = .101$ for speaking, $r = -.282$, $p = .039$ for understanding), monthly use of the second language ($r = .439$, $p = .001$ for speaking, $r = .502$, $p < .001$ for understanding), and language switch frequency (*r* = .421, *p* = .002 for speaking, *r* = .568, *p* < .001 for understanding). Speaking proficiency of the second most proficient language was again used as the measure for Bilingualism, with a score of 0 assigned to participants who listed only one language. There was again no correlation between the measures for WM Capacity and Bilingualism, *r* = .04, *p* = .694. The mean RTs and error rates based on subject means are presented in Table 3.

Table 3

Mean RTs and error rates (and corresponding standard errors) in Experiment 2

	RTs		Error rates	
Congruency	MC items	MI items	MC items	MI items
First block				
Congruent	731 (18)	798 (24)	.019 (.004)	.028 (.008)
Incongruent	919 (24)	846 (21)	.081 (.011)	.044 (.006)
Congruency Effect	188	48	.062	.016
Second block				
Congruent	654 (12)	727 (17)	.017 (.005)	.029 (.010)
Incongruent	844 (22)	802 (19)	.053 (.014)	.043 (.007)
Congruency Effect	190	75	.036	.014

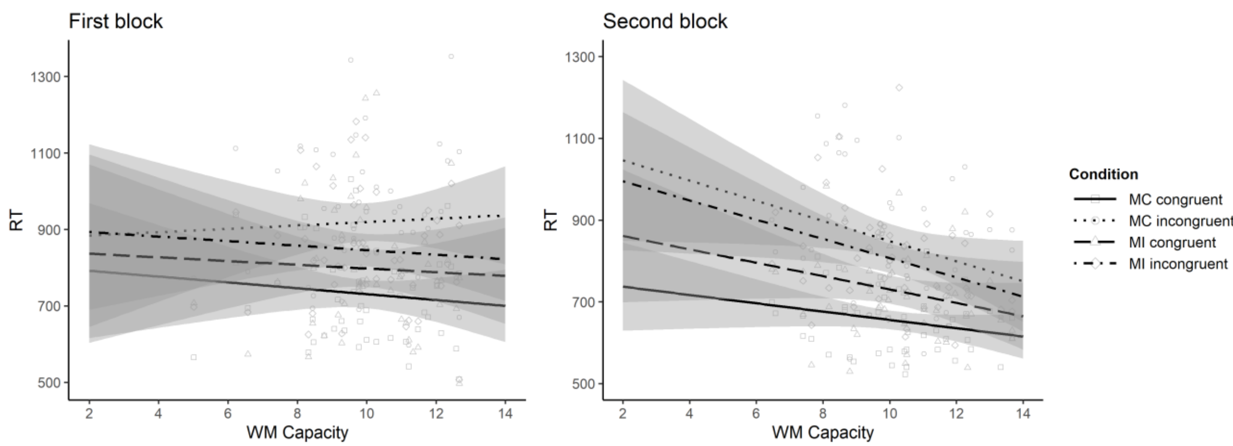
RTs

The results of the latency analyses are reported in Table S4 in the Supplementary Materials. Here, we focus on the most relevant results. As expected, there was an item-specific PCE, with a larger effect of Congruency for MC items than for MI items, $\beta = -28.98$, $SE = 3.28$, $z = -8.84$, $p < .001$, and an effect of Order, with overall latencies decreasing when the item-specific PCE paradigm was presented second rather than first, $\beta = -30.48$, $SE = 4.70$, $z = -6.49$, $p < .001$ (a practice effect; see Table 3). The item-specific PCE was present in both the first and the second block, although it was numerically more pronounced in the first block owing to a rather small congruency effect for MI items in that block. Of primary interest in the analysis were individual differences in the magnitude of the item-specific PCE, which were explored for each order separately in view of the effect of Order. (note 3)

With regards to the item-specific PCE and WM Capacity, Congruency effects were consistently larger for MC items than for MI items, but the item-specific PCE did not vary as a function of WM Capacity in either order. Numerically, however, the item-specific PCE showed an increase with higher WM Capacity in the first block. This result is displayed in Figure 8. The overall effect of Congruency did however vary with WM Capacity, although this was only evident when the item-specific PCE paradigm was administered second: In that situation, not only did higher WM Capacity reduce latencies overall, $\beta = -43.20$, $SE = 8.76$, $z = -4.93$, $p < .001$, but the effect of Congruency also diminished with higher WM Capacity, $\beta = 9.31$, $SE = 4.25$, $z = 2.19$, $p = .028$ (see Figure 9, right panel).

Figure 8

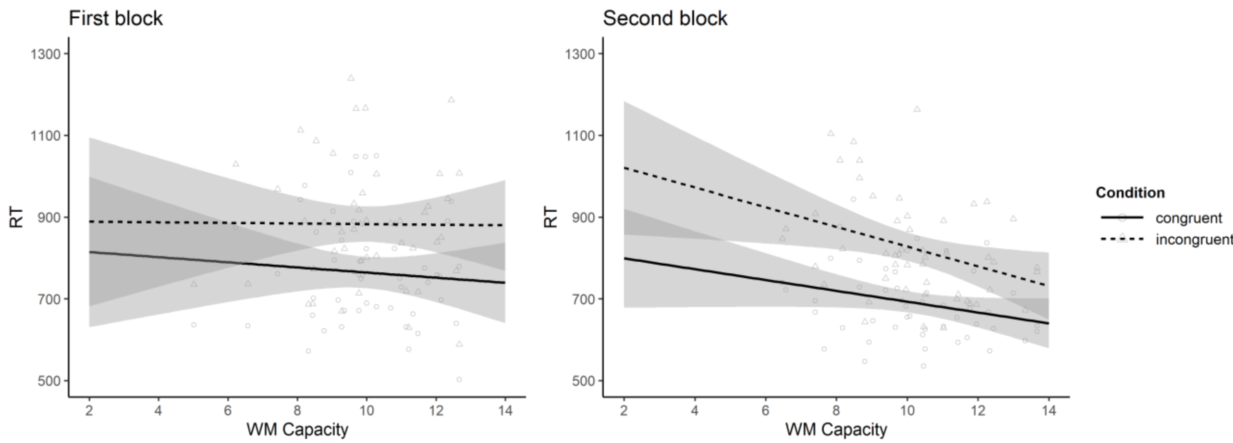
The impact of WM Capacity on the item-specific PCE in latencies in Experiment 2, examined separately for participants who did the Stroop task in the first and second block.



Note. A scatterplot of participants' WM capacity and their mean latencies to congruent items in the MC set (squares), incongruent items in the MC set (circles), congruent items in the MI set (triangles), and incongruent items in the MI set (rhombuses), in the first block (left panel) and second block (right panel). Regression slopes (with 95% confidence interval bands) for congruent items in the MC set, incongruent items in the MC set, congruent items in the MI set, and incongruent items in the MI set, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Figure 9

The impact of WM capacity on the congruency effect in latencies in Experiment 2, examined separately for participants who did the Stroop task in the first and second block.

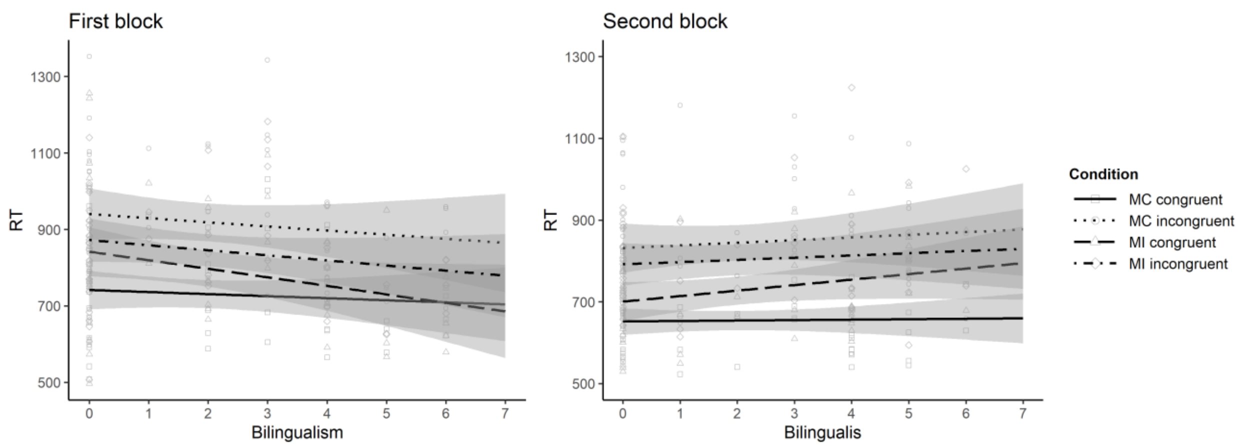


Note. A scatterplot of participants' WM capacity and their mean latencies to congruent (circles) and incongruent (triangles) items, in the first block (left panel) and second block (right panel). Regression slopes (with 95% confidence interval bands) for congruent and incongruent items are marked with a solid line and a dashed line, respectively.

With regards to the item-specific PCE and Bilingualism, there was some evidence that the item-specific PCE increased with increasing Bilingualism, at least for participants who were administered the item-specific PCE paradigm second (see Figure 10, right panel), $\beta = -9.67$, $SE = 4.29$, $z = -2.25$, $p = .024$. In this situation, although higher Bilingualism increased latencies overall, $\beta = 26.23$, $SE = 8.77$, $z = 2.99$, $p = .003$, it did not do so for all conditions equally. For MC items, there was an increase in Congruency effects with increasing Bilingualism driven by a greater slowing to incongruent items, $\beta = 29.99$, $SE = 13.82$, $z = 2.17$, $p = .030$, compared to congruent items, $\beta = 10.72$, $SE = 9.13$, $z = 1.17$, $p = .241$. In contrast, for MI items, there was a decrease in Congruency effects with increasing Bilingualism driven by greater slowing to congruent items, $\beta = 41.70$, $SE = 12.51$, $z = 3.33$, $p < .001$, compared to incongruent items, $\beta = 22.19$, $SE = 9.90$, $z = 2.24$, $p = .025$. Together, these effects contributed to an increase in the item-specific PCE with increasing Bilingualism when the item-specific PCE paradigm was administered second.

Figure 10

The impact of Bilingualism on the item-specific PCE in latencies in Experiment 2, examined separately for participants who did the Stroop task in the first and second block.



Note. A scatterplot of participants' bilingualism scores and their mean latencies to congruent items in the MC set (squares), incongruent items in the MC set (circles), congruent items in the MI set (triangles), and incongruent items in the MI set (rhombuses), in the first block (left panel) and second block (right panel). Regression slopes (with 95% confidence interval bands) for congruent items in the MC set, incongruent items in the MC set, congruent items in the MI set, and incongruent items in the MI set, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

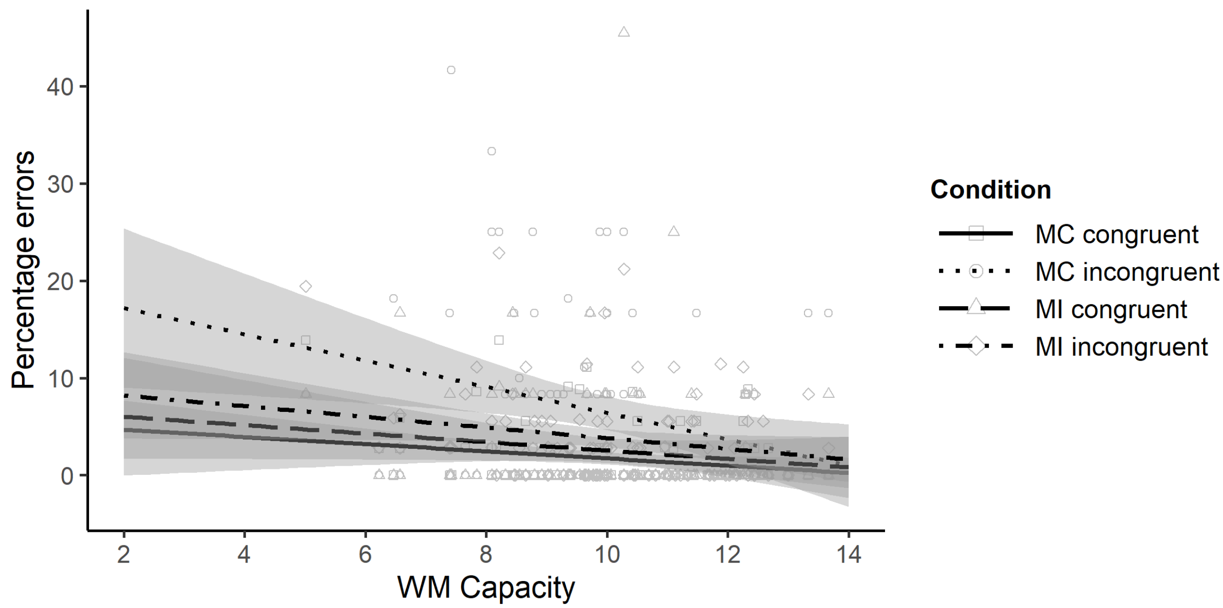
Further evidence that WM capacity and Bilingualism were distinctively related to behaviour came from a 3-way interaction between Bilingualism, WM Capacity, and Order on overall response latency in the main analysis, β = -21.87, *SE* = 4.76, *z* = -4.59, *p* < .001. Follow-up analyses conducted separately for each order revealed that, for participants who were administered the item-specific PCE paradigm in the first block, there was a reduction in overall latencies with increasing WM capacity, but only for participants with low levels of Bilingualism (as determined by a median split on Bilingualism scores), β = -20.16, *SE* = 9.10, *z* = -2.22, *p* = .027. However, for participants who were administered the item-specific PCE paradigm in the second block, there was a reduction in overall latencies with increasing WM capacity that was more pronounced for participants with high levels of Bilingualism, β = -61.30, *SE* = 11.82, *z* = -5.19, *p* < .001, as compared to low levels of Bilingualism, β = -25.43, *SE* = 10.62, *z* = -2.39, *p* = .017.

Errors

The results of the error analyses are reported in Table S5 in the Supplementary Materials. Again, we focus on the most relevant results here. There was a standard item-specific PCE with a larger Congruency effect for MC than MI items, β = .24, *SE* = .08, *z* = 3.08, *p* = .002. This item-specific PCE was the same across both Orders and did not vary with WM Capacity, at least statistically (numerically, there was a tendency for a decrease in the item-specific PCE with higher WM Capacity, see Figure 11), or Bilingualism (Figure 12). Higher WM Capacity reduced errors though, β = .32, *SE* = .12, *z* = 2.69, *p* = .002, whereas Bilingualism did not, β = .00, *SE* = .13, *z* = .04, *p* = .970.

Figure 11

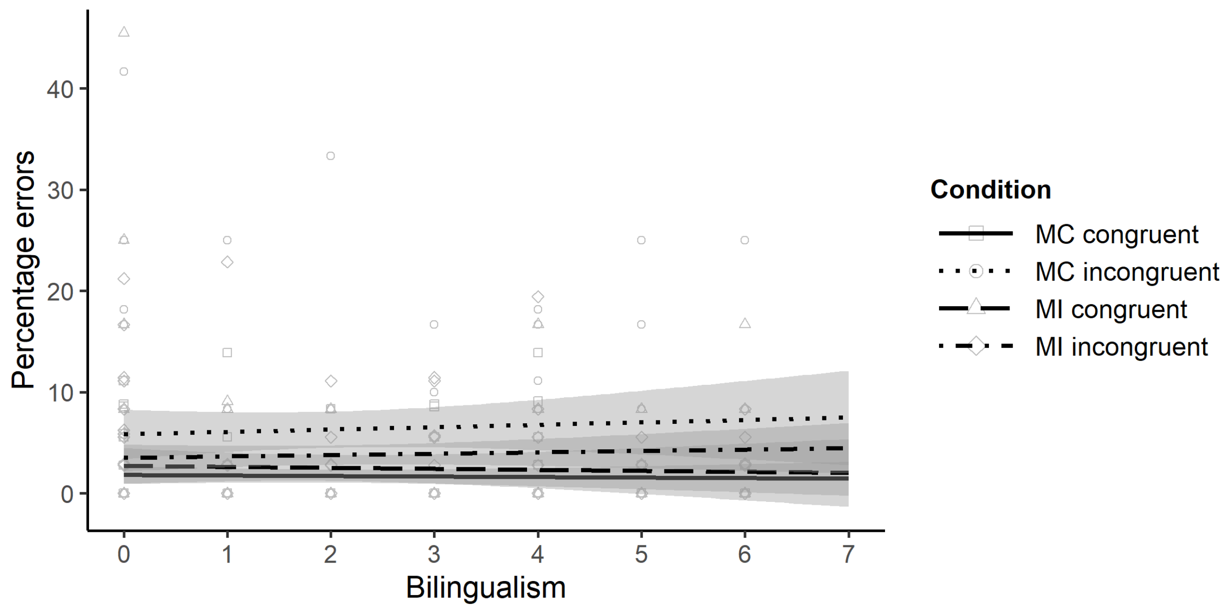
The impact of WM Capacity on the item-specific PCE in error rates in Experiment 2



Note. A scatterplot of participants' WM capacity and their mean error rates for congruent items in the MC set (squares), incongruent items in the MC set (circles), congruent items in the MI set (triangles), and incongruent items in the MI set (rhombuses). Regression slopes (with 95% confidence interval bands) for congruent items in the MC set, incongruent items in the MC set, congruent items in the MI set, and incongruent items in the MI set, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

Figure 12

The impact of Bilingualism on the item-specific PCE in error rates in Experiment 2



Note. A scatterplot of participants' bilingualism scores and their mean error rates for congruent items in the MC set (squares), incongruent items in the MC set (circles), congruent items in the MI set (triangles), and incongruent items in the MI set (rhombuses). Regression slopes (with 95% confidence interval bands) for congruent items in the MC set, incongruent items in the MC set, congruent items in the MI set, and incongruent items in the MI set, are marked with a solid line, a dotted line, a long-dashed line, and a dot-dash patterned line, respectively.

The only other significant effect was a 3-way interaction between Congruency, WM Capacity, and Bilingualism, β = .15, *SE* = .08, z = 1.98, p = .048. This interaction revealed that at low levels of bilingualism (as determined by a median split on Bilingualism scores), congruency effects tended to decrease with increasing WM capacity, β = -.26, *SE* = .18, z = -1.44, p = .150, whereas at high levels of bilingualism, congruency effects tended to increase with increasing WM capacity, β = .18, *SE* = .18, z = 1.04, p = .299.

Discussion

Experiment 2 measured individual differences in reactive control through the use of an item-specific PCE manipulation. Of interest was whether individual differences in reactive control relate to inter-individual variability in WM capacity and language status.

Consistent with previous studies of the item-specific PCE, congruency effects were larger for MC items than MI items for both response latency and accuracy. Thus, participants selected appropriate item-specific control settings (i.e., focused attention for MI items vs. relaxed attention for MC items) without advanced preparation for these items (Bugg & Hutchison, 2013; Spinelli et al., 2020; Spinelli & Lupker, 2020a).

Relations between the item-specific PCE and differences in WM capacity and language status, however, were relatively modest. For example, the item-specific PCE was unrelated to variation in WM capacity. The magnitude of the overall congruency effect did decrease with increasing WM capacity, at least for response times, but there was no indication that the difference in congruency effects for MC and MI items varied in relation to WM capacity. By contrast, the item-specific PCE increased with increasing bilingualism, owing to the fact that increases in bilingualism were associated with progressively slower responses to infrequent items in both the MC and MI conditions (i.e., slower responses to incongruent items in the MC condition and congruent items in the MI condition). Together, these differences yielded an increasing item-specific PCE with increasing bilingualism. Both WM capacity and bilingualism-related effects, however, were only evident when the item-specific PCE paradigm was presented second. Given that participants were randomly assigned to the different order conditions, the most likely explanation is that associations with WM capacity and bilingualism only became evident when participants had become sufficiently proficient with the color-response mappings involved in the paradigm.

Although the effects were somewhat modest, the results of Experiment 2 are informative as they highlight important differences in the way WM capacity and bilingualism relate to attention control. With regards to WM capacity, the findings are overall consistent with the idea that higher WM capacity is associated with a greater likelihood of engaging proactive control (Braver, 2012; Braver et al., 2007; Kane & Engle, 2003), insofar as congruency effects decreased with increasing WM capacity, at least when the item-specific PCE paradigm was administered second. The reason is that, by engaging proactive control, individuals with higher WM capacity are better prepared for conflict than are individuals with lower WM capacity, who mainly rely on reactive control instead.

By contrast, with regards to language status, our results suggest – at least superficially – that greater bilingualism is associated with a greater likelihood of engaging reactive control. When the item-specific PCE paradigm was administered second, higher bilingualism was associated with larger item-specific PCEs. This pattern might reflect the fact that bilinguals rely more on a reactive than a proactive mode of control, allowing them to efficiently select the appropriate control setting (relaxed vs. focused attention) given differences in conflict frequency across MC and MI items. On closer examination, however, variation in the item-specific PCE related to bilingualism is not consistent with the idea of reactive control being the source of that variation. The reason is that changes in congruency effects for MC and MI items associated with higher bilingualism were driven by disproportionate *increases* in latency for *infrequent* incongruent and congruent items in these respective conditions. In contrast, greater reliance on reactive control should mainly result in *decreases* in latency for *frequent* trial types, especially incongruent items in the MI condition for which reactive control permits more efficient conflict resolution. Instead, higher bilingualism was associated with increased latencies in those conditions. Thus, the association between bilingualism and response latencies in the MI and MC conditions was quite idiosyncratic and not easily explained in terms of differences in reactive control.

In general, the results relevant to bilingualism clearly did not parallel those relevant to WM capacity. In fact, there were a few cases in which bilingualism and WM capacity interacted, although they did not do so in a consistent way. In the latencies, WM capacity was more strongly associated with response latencies when bilingualism was high rather than low, at least for participants who did the Stroop task in the second block. This result could potentially be interpreted as showing that bilingualism enhances the improvements afforded by WM capacity on Stroop performance, a form of bilingual advantage. In contrast, for participants who did the Stroop task in the first block, WM capacity had a stronger impact at *lower* than higher levels of bilingualism, a result that may suggest that bilingualism *reduces* the improvements afforded by WM capacity on Stroop performance, a form of bilingual *dis*advantage. Similarly, in the errors, WM capacity had a stronger impact in improving accuracy at higher than lower levels of bilingualism for congruent items, whereas for incongruent items, WM capacity had a stronger impact at *lower* than higher levels of bilingualism. Although this pattern of results is less than clear, what does appear clear is that, overall, bilingualism neither parallels WM capacity in modulating Stroop performance nor does it consistently modulate the impact that WM capacity has on Stroop performance.

To examine these associations more fully, we repeated the above analyses using second-language monthly use and language switch frequency, rather than second-language speaking proficiency, as measures of bilingualism. The results were again similar to the results of the analyses including second-language speaking proficiency as the measure of bilingualism: First, neither second-language monthly use nor language switch frequency correlated with WM capacity, $r = -.05$, $p = .611$, and $r = -.05$, $p = .645$, respectively. Further, while WM capacity modulated overall speed, accuracy, and (in the latencies in the second block) congruency effects, bilingualism did not do so consistently, regardless of how it was measured. Second-language monthly use was associated with higher accuracy, but only for congruent items: As a result, congruency effects in the error rates increased with higher second-language monthly use. Similarly, language switch frequency interacted with congruency in the error rates but the pattern of the interaction was for an increase in the congruency effect with higher language switch frequency.

General Discussion

<u>Language status and cognitive control: Executive attention may not be the right framework</u>

There is much controversy as to whether managing multiple languages affords bilinguals a general advantage over monolinguals in aspects of cognitive control (e.g., Bialystok et al., 2004; Paap & Greenberg, 2013). A recent version of this hypothesis is that this advantage may reside not so much in inhibitory processes, as had been initially thought, but rather in executive attention (Bialystok, 2017) and would thus be more easily detectable in conflict-adaptation paradigms than in simple interference paradigms (Grundy et al., 2017).

Here, we tested this idea using list-wide and item-specific PCE paradigms, gold-standard conflict-adaptation paradigms that generate well-validated measures of proactive and reactive control (Braver, 2012; Braver et al., 2007). Experiment 1 replicated previously reported associations between WM capacity and list-wide PCEs (Hutchison, 2011; Kane & Engle, 2003; Meier & Kane, 2013), such that individuals with a higher WM capacity showed faster and more accurate responses overall, and diminished congruency effects in both MI and MC lists (in which items are mostly incongruent and mostly congruent, respectively). By contrast, Experiment 1 revealed virtually no associations between continuous variation in language status and list-wide PCEs, with the exception of smaller congruency effects associated with increasing bilingualism in the MI list of the MC-first condition. However, that association was quite idiosyncratic, as it was driven by increasing RTs on congruent items rather decreasing RTs on incongruent items. Experiment 2 examined individual differences in item-specific PCEs. Associations with WM capacity were modest and confined to overall decreases in congruency effects with increasing WM capacity in conditions in which color-response mappings were well-practiced. Associations with language status, by contrast, were evident in increases in item-specific PCEs, at least when color-response mappings were well-practiced. However, as in Experiment 1, the pattern of behaviour was quite idiosyncratic, as it was driven largely by longer RTs on infrequent congruent and incongruent items with increasing bilingualism. Thus, across Experiments 1 and 2, continuous variation in language status was associated with what appear to be idiosyncratic differences in conflict adaptation effects, differences that showed little resemblance to differences related to WM capacity.

Note that this general pattern of results remained the same when no participant was removed from the analyses because of their poor accuracy on the distractor component of the complex span tasks (analyses which were based on larger sample sizes, N = 147 for Experiment 1 and N = 123 for Experiment 2, up from N = 125 and N = 98, respectively, in the analyses reported in the main text). In these analyses, list-wide and (to a lesser extent) item-specific PCEs still showed regular associations with WM Capacity but not with Bilingualism, for which those associations, even when diverging from those observed in the main-text analyses, continued to be largely idiosyncratic. For example, different from the main-text analysis, in the error rates of Experiment 1, Bilingualism modified the list-wide PCE in the MC-first condition. However, this modification resulted from the fact that, in that condition, the congruency effect tended to increase in the MI block because Bilingualism improved accuracy for congruent but not incongruent items in that block. Similarly, in the latencies of Experiment 2, Bilingualism was associated with increased latencies overall (although, as in the main-text analysis, this increase was more pronounced in the second block), as well as with a tendency to increase the congruency effect in the first block.

It is also worth noting that for both the sample tested in Experiment 1 and the sample tested in Experiment 2, there was no significant correlation between the measures of language status and WM capacity. This result is consistent with recent meta-analyses which suggest absent or weak associations between bilingualism and WM capacity (Lehtonen et al., 2018; Monnier et al., 2021; but see Grundy & Timmer, 2017). Most importantly, this result is also problematic for the idea that bilingual experience enhances executive attention, as the notion of executive attention has historically been strongly associated with WM capacity (Engle, 2002; Kane et al., 2007). If that idea were true, it would seem to follow that bilingual experience should increase WM capacity. However, there was no evidence for this pattern in our results.

Overall, our findings are inconsistent with the idea that variation in language status is related to differences in executive attention. In interpreting this result, however, a few considerations are in order. One potential drawback of our design is that we used self-rated measures of language proficiency. Self-rated measures are potentially prone to being affected by differences between and within language populations in *perceived* proficiency (Tomoschuk et al., 2019). Because these differences may reduce the interpretability and reliability of self-rated proficiency, objective measures of proficiency (e.g., standardized tests) are arguably more appropriate. That said, self-ratings are widely used both by advocates (e.g., Luk & Bialystok, 2013) and critics (e.g., Paap & Greenberg, 2013) of the bilingual advantage hypothesis, and correlate well with objective measures of language proficiency (Francis & Strobach, 2013; Marian et al., 2007; Paap et al., 2019). Therefore, although objective measures of language proficiency may have provided a more valid assessment of language status, our use of self-rated measures is standard practice in the bilingual advantage literature, especially in studies that use multi-language samples in which receptive vocabulary measurement is typically not feasible. Further, to address concerns about our language proficiency measures, we conducted the analyses again using second-language monthly use and language switch frequency as measures of bilingualism and effectively found the same pattern of results. Thus, our results are robust across measures of language status.

A related issue concerns the fact that, in our analyses, self-rated second-language speaking proficiency was treated as a continuous rather than an ordinal variable (Veríssimo, 2021). Treating self-rated proficiency as ordinal may have been more appropriate, especially if the distribution of values of self-rated speaking proficiency was somewhat unusual. However, that distribution did not appear to be unusual for our participants, and thus it seems unlikely that, in our situation, treating language proficiency as ordinal instead of continuous would have made much of a difference. Further, in our GLMM analyses, an ordinal coding of self-rated proficiency would have increased the complexity of what were already quite complex statistical models, likely leading to convergence failures in model estimation. Finally, although not ideal, the characterization of language status used in the present experiments represents a substantial improvement over the binary characterization of language status used in earlier studies of bilingualism and conflict adaptation effects (e.g., Grundy et al., 2017).

Suffice to say future research is needed to improve the measurement and conceptualization of language status. Gullifer and Titone (2020a, 2020b), for example, suggest that associations between language status and proactive/reactive control may be better captured with the notion of language entropy – the degree to which bilinguals use their languages in a compartmentalized fashion (when they tend to only use one language in a given interactional context) vs. an integrated fashion (when they tend to use multiple languages to a similar extent in a given context). However, this notion (a notion that our background questionnaire was not set up to measure) also requires self-ratings, ratings that are in fact quite complicated (as they require estimating, for several interactional contexts, the percentage of use of each of the languages known by the individual in that context) and that, like self-rated proficiency, may also suffer from problems with interpretability and reliability. Future research will need to establish whether objective measures of second-language proficiency and more advanced subjective measures such as language entropy are able to capture associations between language status and list-wide and item-specific PCEs that resemble those between WM capacity and these effects.

Another potential limitation of our findings concerns the low reliability of latency difference scores and the challenges that this poor reliability poses for the study of individual differences (Draheim et al., 2019; Miller & Ulrich, 2013). This problem is especially true for the congruency effect in the Stroop task (e.g., Hedge et al., 2018) and may have been exacerbated in the present research because PC manipulations inevitably involve reducing the number of observations for the infrequent trial type (e.g., congruent trials in an MI list), as a large number of observations is required for good reliability (Rouder & Haaf, 2019). Consistent with this idea, split-half reliabilities for RT congruency effects (based on means computed from odd- vs. even-numbered occurrences of congruent vs. incongruent trials) as adjusted by the Spearman-Brown prophecy formula were only .662 (MC list) and .238 (MI list) in Experiment 1 and .470 (MC items) and .509 (MI items) in Experiment 2. On the other hand, the type of analyses that we used, i.e., hierarchical models of trial-by-trial performance rather than aggregate performance across trials, has been shown to attenuate these concerns (Rouder & Haaf, 2019). Further, the fact that we replicated previously observed associations between congruency effects and WM capacity suggests that our paradigm was sensitive enough to capture some variation in executive attention, the construct that language status should also be associated with, according to Bialystok (2017). Future research on the relation between language status and executive attention should, however, attempt to use paradigms which are known to afford higher reliability, for example, paradigms that produce accuracy-based rather than RT-based measures of individual abilities (Draheim et al., 2020).

Finally, our conclusions are based on a task, the color-word Stroop task, that involves linguistic mediation. That is, in performing this task, participants necessarily deal with stored linguistic information when retrieving the name of the ink color, matching that color to a response key, and ignoring the name of the color that the word more automatically activates. We chose this task because it is largely with this task that list-wide PCEs, item-specific PCEs, and their associations with WM capacity, have been previously validated (e.g., Hutchison, 2011; Jacoby et al., 2003; Kane & Engle, 2003). However, language status differences in this type of task might largely reflect differences in the domain of language as opposed to differences in domain-general processes such as executive attention, an idea that has led many researchers interested in the bilingual advantage to focus exclusively on non-linguistic tasks (e.g., Bialystok et al., 2009; Hilchey & Klein, 2011; Hilchey et al., 2015; Paap & Greenberg, 2013). On the other hand, it is useful to note that in our experiments, the limited number of linguistic stimuli in the task (four common color names), the fact that manual, not verbal, responses were required, and the fact that the vast majority of our participants were English-dominant and the rest were fluent or near fluent in English (the language of the task) make it unlikely that the linguistic nature of our task could have had a large role in the results. Future research, however, should consider using a non-linguistic task, such as a spatial Stroop task, to strengthen and extend our conclusions.

In any case, as noted, our overall pattern of results is largely consistent with the general idea that variation in WM capacity is positively associated with differences in executive attention (Engle, 2002; Kane et al., 2007). Executive attention primarily favors goal maintenance in the face of distracting information – a proactive process. By constantly maintaining focus on the task goal, individuals with efficient executive attention suffer little conflict not only in contexts that support proactive goal maintenance but also in contexts that do not support it. The result is reduced congruency effects in those individuals, primarily (but not exclusively) in situations that do not support proactive goal maintenance (e.g., Meyer & Kane, 2013). If language status impacts executive attention, as has been proposed (Bialystok, 2017), then language status should have produced comparable differences in conflict adaptation effects. That is not what we found. Therefore, unless the theoretical characterization of executive attention is modified in some way, it does not appear that there is any coherent relationship between language status and executive attention.

Alternatives to executive attention as the locus of the putative bilingual advantage

Although the current evidence is inconsistent with the idea that language status is associated with differences in executive attention, it is also worth considering whether aspects of the present findings are consistent with Grundy et al.'s (2017) suggestion that language status is associated with differences in attentional disengagement. From Grundy et al.'s description, attentional disengagement is a process that terminates attention to preceding stimuli so that preceding stimuli will have little or no impact on current performance. This idea leads to the prediction that in MC lists of list-wide PCE paradigms, greater attentional disengagement should be associated with decreasing congruency effects, since attentional disengagement should nullify the (relaxing) impact of repeated congruent trials. In MI lists, by contrast, greater attentional disengagement should be associated with increasing congruency effects, again because attentional disengagement should nullify the (focusing) impact of repeated incongruent trials. The data from Experiment 1, however, provide no indication that increases in bilingualism were associated with any change in congruency effects in either MC or MI lists, as would be predicted if language status contributed to differences in the efficiency of attentional disengagement.

Another idea as to where the bilingual advantage may be localized in conflict adaptation studies comes from Costa et al. (2009), who compared monolinguals and bilinguals in what they called "low-monitoring" and "high-monitoring" versions of a flanker task. The low-monitoring versions, examined in their Experiment 1, were versions in which the proportion of congruent to incongruent items in the list of trials was either very high (92%, i.e., an MC list) or very low (8%, i.e., an MI list). According to Costa et al., these versions pose little monitoring demands because items are predominantly of one type (either congruent or incongruent) and, thus, participants only infrequently need to switch between items that require conflict resolution (incongruent) and items that do not (congruent). On the other hand, the high-monitoring versions, examined in their Experiment 2, were versions in which the proportion of congruent to incongruent items was more balanced (i.e., 75% in one version and 50% in another version). According to Costa et al., these versions pose more intense monitoring demands because, with the two item types more equally represented, participants need to switch frequently between items that require conflict resolution (incongruent) and items that do not (congruent).

Costa et al. hypothesized that bilingualism affords an advantage not so much in conflict resolution per se but in the monitoring process involved in switching between situations requiring conflict resolution and situations that do not. The implication is that a bilingual advantage should emerge 1) in the form of overall reduced latencies rather than reduced congruency effects and 2) in high-monitoring situations more clearly than in low-monitoring situations. This pattern is indeed the one that they obtained, with no significant differences between monolinguals and bilinguals in the low-monitoring experiments but overall faster latencies for bilinguals in the high-monitoring experiments, especially in the 50% congruent version.

As for attentional disengagement, however, the idea that bilinguals would be advantaged in this particular type of monitoring is difficult to reconcile with the traditional executive attention framework. The main obstacle is that efficient monitoring of this type does not result in reduced congruency effects, the signature pattern of efficient executive attention (but see Arora & Klein, 2020; Grundy, 2020). Relatedly, traditional theories of cognitive control (e.g., Botvinick et al., 2001) assume that what is being monitored is conflict, not item type. In these theories, conflict monitoring activity is higher when conflict is more frequent (e.g., in an MI list), not when item types alternate more frequently (e.g., in a 50% congruent list), making the relevant contrast the one between lists varying in conflict frequency (e.g., MC vs. MI lists), not the one between lists varying in the frequency of item type switches (i.e., the contrast examined by Costa et al., 2009).

Thus, the notion of monitoring as intended by Costa et al. (2009) diverges from that of executive attention in important ways. Further, this notion generates a different set of predictions for the present experiments than those generated by the notions of either executive attention or attentional disengagement. According to Costa et al. (2009), all lists used in the present experiments would be classified as high-monitoring in that they entail frequent switches between items that require conflict resolution (incongruent) and items that do not (congruent). Such would be especially true for the present Experiment 2, which had a congruency proportion of 50% and the highest switch frequency. However, it would also be true for Experiment 1, since congruency proportions of 75% (MC list) and 25% (MI list) still involve moderately frequent switches (remember that Costa et al. classified a 75% congruent list as high-monitoring in their Experiment 2). According to Costa et al. then, because of their superior monitoring abilities, bilinguals should have shown reduced latencies compared to monolinguals in all of those lists, especially in Experiment 2.

As noted, however, there was little evidence to support this idea. In Experiment 1, the only case where bilingualism had an effect on latencies was for congruent items in the MI list in the MC-first condition, and this effect was in the direction of increased rather than decreased latencies. In Experiment 2, while higher bilingualism did tend to speed up latencies when the item-specific PCE paradigm was presented first (especially for MI items; see the left panel of Figure 10), it also tended to slow down latencies when the item-specific PCE paradigm was presented second (the situation where bilingualism also altered the item-specific PCE). Thus, the idea that the bilingual advantage may lie in more efficient monitoring processes facilitating switches between different item types also seems to make little headway in explaining the present data.

A final alternative that may be considered as a potential locus of the bilingual advantage is one that is entirely based on the obtained data pattern. The main pattern associated with bilingualism in the present experiments was the increased item-specific PCE in the second block of Experiment 2. Insofar as the item-specific PCE is thought to reflect primarily reactive control (e.g., Spinelli et al., 2020), this pattern may potentially reflect a tendency among bilinguals to rely more on a reactive than a proactive mode of control. In this view, bilinguals may not be particularly advantaged in processes pertaining to proactive control, the control mode typically associated with efficient executive attention. Instead, they would be efficient in reactive processes, processes that, in an item-specific PCE paradigm, permit adoption of control settings that are appropriate to the specific item and would thus result in a pronounced item-specific PCE. This idea appears to make some intuitive sense, as bilinguals certainly need to rely on cues in the environment to activate the context-appropriate language and regular practice with doing so may enhance reactive control abilities.

On the other hand, it must be noted that the increased item-specific PCE in bilinguals was driven by variability in the latency of the infrequent trial types (rather than the frequent trial types, as this idea would predict), and it was only observed in one situation (i.e., when the item-specific PCE was presented second), a situation in which there was a general slowdown associated with bilingualism. Such a slowdown would seem to be difficult to reconcile with the idea of a bilingual advantage. This pattern is thus potentially idiosyncratic and further research is required to ascertain its robustness.

Conclusions

The idea recently proposed by Bialystok (2017) that bilingualism confers an advantage in executive attention found no support in the present research. More efficient executive attention, as indexed by WM capacity, was linked to greater utilization of proactive control, but bilingual language status was not. Interpretations that assume a different locus for the bilingual advantage than executive attention, such as attentional disengagement and monitoring, also fell short in terms of explaining the present data. Although some evidence did emerge suggesting that bilinguals rely more on a reactive than proactive mode of control, this evidence is not without problems either and requires replication. In summary then, the present data challenge all current conceptualizations of the bilingual advantage and call for yet another revision of that hypothesis.

Notes

1. In the errors, we did get evidence for WM capacity to differentially affect congruency effects in MC vs. MI blocks, specifically, congruency effects decreased with higher WM capacity in MI blocks but not in MC blocks. However, we do not think that this pattern is particularly meaningful. The reason is that this pattern originated from the fact that congruent items in MI blocks were the only items for which accuracy did not improve with higher WM capacity. Because accuracy for those items was likely at ceiling (average error rate was 1.6%), observing an improvement in accuracy for those items may not have been possible.
2. In the errors, another situation in which there was a tendency for congruency effects to be reduced with bilingualism was the MI-first condition. However, in this situation as well, the reduction was mainly driven by decreased accuracy on congruent items rather than increased accuracy on incongruent items. As noted, such a pattern prevents us from drawing conclusions about the efficiency with which individuals resolve interference. A similar point could be made about the five-way interaction that we found in the errors, the finding that in the MI block in the MI-first condition specifically, WM capacity led to worse accuracy on congruent items at lower levels of bilingualism. This pattern was, once again, driven by congruent items, items that are not crucial for the question of how individuals deal with conflict.
3. The model for participants who did the non-conflict task first failed to converge. As per the recommended troubleshooting procedure (see “convergence” help page in R), we restarted the model from the apparent optimum and tried other optimizers in addition to the BOBYQA optimizer. Although all optimizers failed to converge, all optimizers except for BOBYQA returned very similar results, suggesting that the convergence warnings were false positives. We report the results from lme4’s default optimizer for GLMMs.

Author note

This research was partially supported by an NSERC Discovery Grant and a SSHRC Insight Grant R3353 to J. Bruce Morton, an NSERC Discovery Grant A633 to Stephen J. Lupker, and an Ontario Trillium Scholarship to Giacomo Spinelli. We wish to thank Nicole Takken and Kesheni Krishna for their help in data collection for Experiments 1 and 2, respectively, and Ken Paap for sharing his language-use questionnaire with us. The raw data and R and SPSS files used for the analyses are publicly available at https://osf.io/gz6h3/. The study materials are available upon request. The studies were not preregistered.

References

Abrahamse, E. L., Duthoo, W., Notebaert, W., & Risko, E. F. (2013). Attention modulation by proportion congruency: The asymmetrical list shifting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1552–1562. https://doi.org/10.1037/a0032426

Arora, S. & Klein, R. M. (2020 Comparing bilingual and monolingual performance on the Attention Network Test: Meta-analysis of a literature inspired by Albert Costa. *Journal of Cultural Cognitive Science*, *4*, 243-257. https://doi.org/10.1007/s41809-020-00068-z

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48. https://doi.org/10.18637/jss.v067.i01

Bialystok, E. (2017). The bilingual adaptation: how minds accommodate experience. *Psychological Bulletin*, *143*, 233-262. https://doi.org/10.1037/bul0000099

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*, 89-129. https://doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, Aging, and Cognitive Control: Evidence From the Simon Task. *Psychology and Aging*, *19*, 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 859-873. https://doi.org/10.1037/0278-7393.34.4.859

Bolker, B. (2020). *GLMM FAQ*. Retrieved from https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624-652. https://doi.org/10.1037/0033-295X.108.3.624

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A.R.A. Conway, C. Jarrold, M.J. Kane, A. Miyake, & J.N. Towse (Eds.), *Variation in working memory* (pp. 76-108). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0004

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106-113. https://doi.org/10.1016/j.tics.2011.12.010

Bugg, J. M., & Hutchison, K. A. (2013). Converging evidence for control of color–word Stroop interference at the item level. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 433-449. https://doi.org/10.1037/a0029145

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769-786. https://doi.org/10.3758/bf03196772

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*, 135-149. https://doi.org/10.1016/j.cognition.2009.08.001

De Bruin, A., Bak, T. H., & Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *Journal of Memory and Language*, *85*, 15-26. https://doi.org/10.1016/j.jml.2015.07.001

Đokić, R., Koso-Drljević, M., & Đapo, N. (2018). Working memory span tasks: Group administration and omitting accuracy criterion do not change metric characteristics. *PloS one*, *13*, e0205169.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*, 508-535. https://doi.org/10.1037/bul0000192

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2020). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*, 242-275. https://doi.org/10.1037/xge0000783

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited. *Experimental psychology*, *61*, 234-251. https://doi.org/10.1027/1618-3169/a000243

Egner, T. (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology*, *5*, 1247. https://doi.org/10.3389/fpsyg.2014.01247

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19-23. https://doi.org/10.1111/1467-8721.00160

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143-149. https://doi.org/10.3758/BF03203267

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116-124. https://doi.org/10.3758/BF03195503

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*, 226-236. https://doi.org/10.3758/s13421-014-0461-7

Francis, W.S., & Strobach, E. N. (2013, November). *Prediction of objective bilingual language proficiency: Can self-report measures suffice?* [Conference presentation] 54th Annual Meeting of the Psychonomic Society, Toronto, Canada.

Goldsmith, S. F., & Morton, J. B. (2018a). Sequential congruency effects in monolingual and bilingual adults: a failure to replicate Grundy et al. (2017). *Frontiers in Psychology*, *9*, 2476. https://doi.org/10.3389/fpsyg.2018.02476

Goldsmith, S. F., & Morton, J. B. (2018b). Time to disengage from the bilingual advantage hypothesis. Cognition, 170, 328–329. https://doi.org/10.1016/j.cognition.2017.09.011

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*, 778-788. https://doi.org/10.3758/s13421-016-0591-1

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

Grundy, J. G. (2020). The effects of bilingualism on executive functions: An updated quantitative analysis. *Journal of Cultural Cognitive Science*, *4*, 177–199. https://doi.org/10.1007/s41809-020-00062-5

Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L., & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, *163*, 42-55. https://doi.org/10.1016/j.cognition.2017.02.010

Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, *33*, 325-340. https://doi.org/10.1177/0267658316678286

Gullifer, J. W., & Titone, D. (2020a). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*, 283-294. Advance online publication. https://doi.org/10.1017/S1366728919000026

Gullifer, J. W., & Titone, D. (2020b). Engaging proactive control: Influences of diverse language experiences using insights from machine learning. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0000933

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166-1186. https://doi.org/10.3758/s13428-017-0935-1

Hilchey, M. D. & Klein, R. M. (2011) Are there bilingual advantages on non-linguistic interference tasks? Implications for plasticity of executive control processes, *Psychonomic Bulletin & Review*, *18*, 625-658. https://doi.org/10.3758/s13423-011-0116-7

Hilchey, M. D., Saint-Aubin, J. & Klein, R. M. (2015) Does bilingual exercise enhance cognitive fitness in traditional non-linguistic executive processing tasks? In J. Schwieter (Ed.), *The Cambridge Handbook of Bilingual Processing* (pp. 586-613). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781107447257.026

Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 851-860. https://doi.org/10.1037/a0023437

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*, 638-644. https://doi.org/10.3758/BF03196526

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A.R.A. Conway, C. Jarrold, M.J. Kane, A. Miyake, & J.N. Towse (Eds.), *Variation in working memory* (pp. 21-48). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0002

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70. https://doi.org/10.1037/0096-3445.132.1.47

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-271. https://doi.org/10.1037/0096-3445.133.2.189

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

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. Retrieved from https://CRAN.R-project.org/package=emmeans

Levy, R. (2014). Using R formulae to test for main effects in the presence of higher order interactions. *arXiv:1405.2094*.

Lin, O. Y.-H., & MacLeod, C. M. (2018). The acquisition of simple associations as observed in color–word contingency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 99–106. https://doi.org/10.1037/xlm0000436

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. Chicago: Psychology Press. https://doi.org/10.4324/9781410609823

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*, 166-174. https://doi.org/10.3758/BF03197535

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*, 605-621. https://doi.org/10.1080/20445911.2013.795574

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940-957. https://doi.org/10.1044/1092-4388(2007/067)

Meier, M. E., & Kane, M. J. (2013). Working memory capacity and Stroop interference: Global versus local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 748-759. http://dx.doi.org/10.1037/a0029200

Meier, M. E., & Kane, M. J. (2015). Carving executive control at its joints: Working memory capacity predicts stimulus–stimulus, but not stimulus–response, conflict. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1849-1872. http://dx.doi.org/10.1037/xlm0000147

Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review*, *20*, 819-858. https://doi.org/10.3758/s13423-013-0404-5

Mishra, R. K., Hilchey, M. D., Singh, N. & Klein, R. M. (2012) On the time course of exogenous cuing effects in bilinguals: Higher proficiency in a second language is associated with more rapid endogenous disengagement, *Quarterly Journal of Experimental Psychology*, *65*, 1502-1510. https://doi.org/10.1080/17470218.2012.657656

Monnier, C., Boiché, J., Armandon, P., Baudoin, S., & Bellocchi, S. (2021). Is bilingualism associated with better working memory capacity? A meta-analysis. *International Journal of Bilingual Education and Bilingualism*. Advance online publication.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232-258. https://doi.org/10.1016/j.cogpsych.2012.12.002

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

Paap, K. R., Myuz, H., Anders-Jefferson, R., Mason, L., & Zimiga, B. (2019). On the ambiguity regarding the relationship between sequential congruency effects, bilingual advantages in cognitive control, and the disengagement of attention. *AIMS Neuroscience*, *6*, 282-298. https://dx.doi.org/10.3934%2FNeuroscience.2019.4.282

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

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*, 164-171. https://doi.org/10.1027/1015-5759/a000123

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*, 452-467. https://doi.org/10.3758/s13423-018-1558-y

Saint-Aubin, J., Hilchey, M. D., Mishra, R., Singh, N., Savoie, D., Guitard, D., & Klein, R. M. (2018). Does the relation between the control of attention and second language proficiency generalize from India to Canada? *Canadian Journal of Experimental Psychology*, *73*, 208-218. https://doi.org/10.1037/cep0000151

Schmidt, J. R. (2013). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*, *20*, 615-630. https://doi.org/10.3758/s13423-012-0373-0

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review*, *26*, 753-771. https://doi.org/10.3758/s13423-018-1520-z

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 514–523. https://doi.org/10.1037/0278-7393.34.3.514

Schmidt, J. R., Crump, M. J., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*, 421-435. https://doi.org/10.1016/j.concog.2006.06.010

Spinelli, G., Krishna, K., Perry, J. R., & Lupker, S. J. (2020). Working memory load dissociates contingency learning and item-specific proportion-congruent effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 2007–2033. https://doi.org/10.1037/xlm0000934

Spinelli, G., & Lupker, S. J. (2020a). Item-specific control of attention in the Stroop task: Contingency learning is not the whole story in the item-specific proportion-congruent effect. *Memory & Cognition*, *48*, 426–435. https://doi.org/10.3758/s13421-019-00980-y

Spinelli, G., & Lupker, S. J. (2020b). Proactive control in the Stroop task: A conflict-frequency manipulation free of item-specific, contingency-learning, and color-word correlation confounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0000820

Spinelli, G., Perry, J. R., & Lupker, S. J. (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture–word interference task. *Journal of Experimental Psychology: Human Perception and Performance*, *45*, 995–1014. https://doi.org/10.1037/xhp0000656

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662. https://doi.org/10.1037/h0054651

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, *22*, 516-536. https://doi.org/10.1017/S1366728918000421

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language*, *28*, 127-154. https://doi.org/10.1016/0749-596x(89)90040-5

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505. https://doi.org/10.3758/bf03192720

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635-654.

Van Steenbergen, H., Haasnoot, E., Bocanegra, B. R., Berretty, E. W., & Hommel, B. (2015). Practice explains abolished behavioural adaptation after human dorsal anterior cingulate cortex lesions. *Scientific Reports*, *5*, 9721.

Veríssimo, J. (2021). *Analysis of rating scales: A pervasive problem in bilingualism research and a solution with Bayesian ordinal models*. PsyArXiv. https://doi.org/10.31234/osf.io/4m2ap

Verreyt, N., Woumans, E. V. Y., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, *19*, 181-190. https://doi.org/10.1017/S1366728914000352

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. New York: Springer. https://doi.org/10.1007/978-3-319-24277-4

Yang, H., Chen, J., Spinelli, G., & Lupker, S. J. (2019). The impact of text orientation on form priming effects in four-character Chinese words. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 45*, 1511-1526. https://doi.org/10.1037/xlm0000655