The Inhibitory Advantage in Bilingual Children Revisited

Myth or Reality?

Jon Andoni Duñabeitia,[1] Juan Andrés Hernández,[2] Eneko Antón,[1] Pedro Macizo,[3] Adelina Estévez,[2] Luis J. Fuentes,[4] and Manuel Carreiras[1,5,6]

[1]Basque Center on Cognition, Brain and Language (BCBL), Donostia, Spain, [2]Universidad de La Laguna, Tenerife, Spain, [3]Universidad de Granada, Spain, [4]Universidad de Murcia, Spain, [5]Ikerbasque, Basque Foundation for Science, Bilbao, Spain, [6]Universidad del País Vasco (UPV/EHU), Bilbao, Spain

Abstract. In recent decades several authors have suggested that bilinguals exhibit enhanced cognitive control as compared to monolinguals and some proposals suggest that this main difference between monolinguals and bilinguals is related to bilinguals' enhanced capacity of inhibiting irrelevant information. This has led to the proposal of the so-called bilingual advantage in inhibitory skills. However, recent studies have cast some doubt on the locus and generality of the alleged bilingual advantage in inhibitory skills. In the current study we investigated inhibitory skills in a large sample of 252 monolingual and 252 bilingual children who were carefully matched on a large number of indices. We tested their performance in a verbal Stroop task and in a nonverbal version of the same task (the number size-congruency task). Results were unequivocal and showed that bilingual and monolingual participants performed equally in these two tasks across all the indices or markers of inhibitory skills explored. Furthermore, the lack of differences between monolingual and bilingual children extended to all the age ranges tested and was not modulated by any of the independent factors investigated. In light of these results, we conclude that bilingual children do not exhibit any specific advantage in simple inhibitory tasks as compared to monolinguals.

Keywords: bilingual advantage, inhibitory skills, executive control, Stroop effects

One important cognitive skill is the ability to ignore irrelevant information in a given situation and to focus on what is needed. This prevents irrelevant investment of cognitive resources. Obviously, this skill is not equally developed in all human beings. One of the most intriguing examples of between-group differences in inhibitory skills is provided by the comparison between monolinguals and bilinguals. Bilingual speakers have to negotiate two languages in everyday situations. They are aware of the linguistic context and of the listeners' language in conversations and they have to focus on the language they want to use, inhibiting the nonrelevant language.

Early experimental reports suggested that bilinguals have enhanced executive control as compared to their monolingual peers, given the apparent advantages of bilinguals in classical markers associated with inhibitory skills (e.g., Stroop or Simon effects, among others; see Bialystok, Craik, & Luk, 2012, for review; see also Bialystok, 2009; Bialystok, Craik, & Luk, 2008; Colzato et al., 2008). As recently reviewed by Bialystok et al. (2012), there are (at least) two main theoretical views regarding the origin of this "bilingual advantage." On the one hand, according to some proposals, it is suggested that bilinguals' enhanced training in inhibiting the irrelevant information (e.g., the nonrelevant language) provides them with better inhibitory capacities than monolinguals (see Bialystok, Craik, Klein, & Viswanathan, 2004; Green, 1998; Kroll, Bobb, Misra, & Guo, 2008). Bilinguals' continuous need to inhibit the nontarget language stems from the fact that both languages (the L1 and the L2 in nonbalanced bilinguals, or the multiple L1s in the case of simultaneous balanced bilinguals) are always active, even in a pure monolingual context where one of the languages is not even mentioned or present (e.g., Thierry & Wu, 2007). On the other hand, alternative views suggest that the bilingual advantage in executive control does not directly respond to enhanced inhibitory mechanisms, but rather to enhanced conflict-monitoring skills and goal orienting (e.g., Colzato et al., 2008; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Hernández, Costa, Fuentes, Vivas, & Sebastián-Gallés, 2010; see Bialystok et al., 2012, and Hilchey & Klein, 2011, for review). Taking into account the above results, the present study focuses on the extent to which differences in inhibitory mechanisms could underlie the so-called bilingual advantage in executive control.

One of the most popular and most widely-studied tasks that has been used to measure inhibitory control is the Stroop task (Stroop, 1935). According to the general claim that bilinguals outperform monolinguals in tasks that require inhibition of irrelevant information, it could be expected that bilinguals perform much better than monolinguals in this task. In fact, this has been a finding often reported in recent years. Bialystok and collaborators tested old (around 68 years old) and young bilinguals (around 20 years old) in a Stroop task (Bialystok et al., 2008). Participants were required to name the color in which the displayed words were printed, without paying attention to the actual words themselves. Words could be congruent with the color they were printed in (e.g., the word "green" printed in green ink) or incongruent (e.g., the word "green" printed in red ink). Together with these critical conditions, other control stimuli were interspersed: A control symbol condition (a sequence of Xs in one of the target colors) and a control word condition (words printed in black ink, for which participants had to read the words aloud). The authors found no differences between the language groups when they analyzed the congruent and incongruent trials together (namely, the overall reaction times across conditions), but when they computed the differences in the reaction times (RTs) for congruent and the incongruent trials (i.e., the Stroop effect), they found that this effect interacted with both age and language groups: They found a larger Stroop effect for older monolinguals as compared to their bilingual peers. Furthermore, when they analyzed the facilitation effects (i.e., RTs in congruent trials minus RTs in control symbol trials) and the interference effects (i.e., RTs in incongruent vs. control trials), they found that the facilitation effect was larger and that the interference effect was smaller in the bilingual sample than in monolingual sample (see Hernández et al., 2010, for a similar pattern). However, as stated, these differences were modulated by the age of the participants and were mainly present in the older group of participants. This is somewhat parallel to other findings showing that the impact of bilingualism on nonlinguistic inhibitory skills is primarily evident in advanced stages of life (e.g., Bialystok, Craik, & Freedman, 2007; see Hilchey & Klein, 2011, for review).

How conflicting information is managed in monolinguals and bilinguals and how this skill is affected by age have been also examined using the Simon task (Simon & Rudell, 1967). In this task participants are asked to respond according to one dimension or feature of the stimuli (e.g., the color) by providing responses with the right and left hands, while ignoring some other dimensions of the same stimuli (e.g., their position on the screen) that represent either a congruency or a conflict with respect to the hand with which they should respond. When bilinguals and monolinguals perform this task, the Simon effect (i.e., longer latencies for incongruent than for congruent trials) has been shown to be smaller for bilinguals than for monolinguals (e.g., Bialystok et al., 2004), and this between-group difference is most clearly found in older participants, whereas in younger adults the effect is not so strong. (Note, however, that the number of studies testing inhibitory control in old bilinguals is remarkably low and that the pattern of composite scores from which the Simon effect is derived is somewhat heterogeneous (e.g., Bialystok et al., 2008)). In these and other studies testing young samples, Bialystok and collaborators (among others) have found that the differential effects in dual-task processing between monolinguals and bilinguals are typically small (e.g., Bialystok, 2006; Bialystok, Craik, & Ruocco, 2006). Hence, we wanted to explore if the advantage of bilinguals over monolinguals in tasks involving inhibitory control would be evident in children of different ages. Our guiding hypothesis is that the "bilingual advantage" in inhibitory skills could be remarkably instable across the lifespan, leading to small differential effects in young adults as compared to older adults. As we will suggest in the Discussion, it could be considered that the bilingual advantage on inhibitory control is mainly evident in behavioral measures obtained from samples of older participants that are not in their peak of cognitive functioning and that it is not stably manifested in younger samples (e.g., young adults or children).

In the present study we explored the performance of a large sample of bilingual children of different ages from a bilingual community and of matched monolingual children from monolingual environments in two versions of the Stroop task with different linguistic demands. It is important to be aware of an inherent problem of the Stroop task when dealing with language-based test groups: In this task it is admittedly difficult to isolate differences in pure inhibitory skills from differences mainly due to basic linguistic performance variations which are linked to participants' linguistic skills and proficiency, given that the classical version of the task necessarily involves linguistic stimuli. For this reason, researchers investigating the relationship between executive control and multilingualism have recently adopted different approaches to systematic investigation of the differential executive function-related effects in monolinguals and bilinguals, using nonlinguistic tasks. The Attention Network Test (ANT; see Fan, McCandliss, Sommer, Raz, & Posner, 2002), a task tapping into executive control as well as into alerting and orienting networks, has been successfully used in this regard. In this task, participants have to decide in which direction an arrow is pointing. The arrow can have congruent arrows at its sides (pointing in the same direction) or incongruent arrows (pointing in the opposite direction). In addition, a cue can be presented either in the same place as the flanker arrows are going to be presented, or in another location. Costa, Hernández, and Sebastián-Gallés (2008) tested 200 young Catalan-Spanish bilinguals (mean age: 22 years) with the ANT task. They found that, regardless of the type of trial (congruent, incongruent, or neutral), bilingual participants were faster than monolinguals in their general performance. Crucially, the conflict effect (incongruent vs. congruent trials) was found to be different for monolingual and bilinguals, with this effect significantly larger for monolinguals than for bilinguals. Nevertheless, this differential effect disappeared during the course of the experiment, probably as a function of within-task specialization or adaptation strategies. This study (and other closely related findings; e.g., Bialystok & Martin, 2004; Carlson & Meltzoff, 2008) suggest that bilinguals are overall faster in tasks that involve conflict resolution and that the interference effect produced by the incongruent trials is larger for monolinguals than for bilinguals.

However, the general picture of bilingual advantage in the management of conflicting information is not straightforward. In a series of experiments exploring the influence of bilingualism in inhibitory skills of young adults, Paap and Greenberg (2013) demonstrated that the cross-study reliability and replicability of the bilingual advantage effect is markedly low. In their study, Paap and Greenberg collected data from a large group of bilinguals who completed a series of Simon, flanker, and antisaccade tasks and the authors did not detect any evidence of bilingual advantage in these tasks that inherently involve inhibitory mechanisms and skills. They reviewed previous evidence and including their three reports identified a total of 18 tests yielding no bilingual advantage, concluding that previous studies reporting such advantage are very likely to be a consequence of Type I error, inadequately matched groups, uncontrolled external factors, and/or task-dependent effects (see also Morton & Harper, 2007).

In a different vein, Costa et al. (2009) also noticed that the magnitude of the conflict effects and the overall RTs between groups were highly similar in a large proportion of the studies reported in the literature (e.g., Bialystok, 2006; Bialystok et al., 2008), and in cases in which a bilingual advantage is found, it is most likely to be present in the form of an overall RT difference between groups, rather than in the magnitude of the interference/facilitation effects (Bialystok, Martin, & Viswanathan, 2005; Morton & Harper, 2007; see also Hernández et al., 2010, for a general bilingual difference across all conditions in a Stroop task). Costa et al. demonstrated that while the overall RT bilingual advantage is relatively reliable in adult samples (namely, shorter RTs for bilinguals than for monolinguals in the ANT), the differences in effect sizes are inconsistent and highly dependent on strategic factors that may arise during the course of the experiments. The explanation provided by Costa et al. for the bilingual advantage in the general task performance and for its absence in the individual effects associated with the different components of the attentional network was that the advantage stems from the impact of conflict monitoring, rather than from the inhibitory capacity per se. According to the monitoring explanation, bilinguals outperform monolinguals in cognitive flexibility mechanisms that allow them to change between tasks or trials that have different requirements (e.g., from conditions requiring conflict resolution to conditions that do not require so), similarly to the way they change from one language to another depending on the context (i.e., language switching). This explanation also readily explains the fact that the overall RT bilingual advantage is found in mixed-design experiments rather than in block-design experiments, since in the latter there is no need for adaptation from one trial to the next one because participants can infer the within-block consistency. Still, the sample tested was composed of young adults and the extent to which the bilingual advantage is present (or absent) in younger samples remains to be seen. Whether or not global RTs represent an accurate measure of monitoring or goal-orientation is a matter of debate, given that in clear contrast to other classical indices in the literature on executive control that represent difference scores between conditions (e.g., the Simon or Stroop effects), raw response times are markedly influenced by individual differences in basic motor and perceptual processes. Moreover, it is worth noting that the so-called bilingual advantage in general task performance (global RTs) is a sometimes elusive phenomenon, and that several authors have failed at showing shorter RTs for bilinguals than for monolinguals in a variety of tasks including Stroop, Simon, and flanker tasks (e.g., Kousaie & Phillips, 2012; see Paap & Greenberg, 2013, for review).

Given the need for methodical investigation of the bilingual advantage, in general, and of the differences between monolinguals and bilinguals in inhibitory skills, in particular, we explored this issue in two Stroop experiments with carefully-selected large samples of more than 250 bilingual and 250 monolingual children of different ages recruited from different elementary and high schools. Considering the difficulties in teasing apart purely linguistic factors from factors associated with inhibitory control in the classic Stroop task, given the direct involvement of language in the task, we followed Hernández et al. (2010) and tested participants with a task that clearly implied less explicit activation of linguistic knowledge, but that also assesses the management of conflicting information: The numerical Stroop task. This task, also called the number-size congruency task (see Kadosh, Gevers, & Notebaert, 2011; see also Jolicoeur, 1987), requires that participants decide which of two visually displayed digits is bigger in size than the other, without paying attention to the numerical magnitude represented by each of those digits (i.e., inhibiting their numeric meaning). Following the conditions typically implemented in the classical linguistic Stroop task, the numerical Stroop task also involves congruent situations (e.g., a small 3 vs. a big 7), incongruent situations (e.g., a big 3 vs. a small 7), and neutral situations (e.g., a small 3 vs. a big 3). Considering that bilinguals and monolinguals regularly display differences in the time needed to complete lexical access (e.g., Bialystok et al., 2008; Ivanova & Costa, 2008), this task will allow exploration of the inhibitory ability of the participants in a scenario requiring less direct explicit implication of the linguistic system than the classic Stroop task. Therefore, we conducted two Stroop experiments with clearly different linguistic demands, in order to compare the potential advantage of bilinguals over monolinguals in tasks involving inhibitory control. Comparing the results of the two versions of the Stroop task provides an excellent opportunity to investigate how language-related enhanced inhibitory capacities (e.g., the bilingual advantage in inhibitory control) develop over the course of schooling and the extent to which these effects are modulated as a function of increased involvement of language in the tasks in hand. If bilinguals have better inhibitory skills than monolinguals, we should then observe reduced inhibitory effects in the Stroop tasks in bilinguals compared to monolinguals. Besides, as suggested by the general conflict-monitoring hypothesis (e.g., Bialystok et al., 2004; Costa et al., 2009; see Hilchey & Klein, 2011), global reaction time differences could also be predicted between the samples, with bilinguals faster overall than monolinguals in the tasks. However, taking into account recent evidence against the bilingual advantage in simple inhibitory tasks (see Paap & Greenberg, 2013), unambiguous between-group differences may not be observed.

Experiment 1: Classic Stroop Task

Method

Participants

Two groups of participants were recruited from different elementary and high schools in Spain ($n = 504$; 280 females). The first group was made up of 252 Spanish monolingual children (137 females) from Grades 3, 4, 5, and 6 of different elementary schools and from Grades 1 and 2 of different high schools (42 monolingual participants from each grade). The mean age of each grade can be found in Table 1. These Spanish-speaking monolinguals had no fluent knowledge of any other language and were recruited at different schools from Spanish provinces where Spanish is the only official language. None of them corresponded to any immigrant minority. Furthermore, we made sure that they were exclusively exposed to Spanish at home through a questionnaire that was completed by their parents or legal tutors. The second group was made up of 252 Basque-Spanish bilingual children (143 females) from the same grades as the monolingual children (42 bilingual participants from each grade). All these bilingual participants were recruited from schools in the Basque Country, a Spanish region where Spanish and Basque are co-official languages that are present in everyday life. All the bilingual children were attending bilingual schools where the two languages were being used as vehicular languages in the educational practice (bilingual linguistic educational model). This linguistic model is based on a legal regulation that ensures that students are exposed to the two languages at school in an active manner, switching languages between the different academic subjects. Thus academic subjects are distributed following a ratio of 50% in each language (Basque and Spanish). We carefully selected the bilingual participants to ensure that all of them were born in the Basque Country and that none of them corresponded to any specific social minority. Bilingual and monolingual participants were carefully matched in different measures and cognitive skills (see Table 1 for detailed information).

Except in the case of 1 bilingual participant, the parents of all the participants in the study reported Spanish as their first language (L1). (Note in this regard that some of the parents of the bilingual group also reported knowledge of Basque, but did not report this as their L1). A linguistic-competence questionnaire completed by the bilingual children's parents (fully available for 241 out of the 252 bilingual children) showed that bilingual participants had acquired Spanish earlier in life than Basque (Spanish AoA, in years: mean = 0.75, $SD = 0.89$; Basque AoA: mean = 2.27, $SD = 1.11$), and that taking into account the nonacademic context, they were more exposed to Spanish than to Basque (percentage of time exposed to Spanish: mean = 65.14%, $SD = 13.42\%$). Their mean competence level in Spanish on a 10-point scale was 8.68 ($SD = 1.23$), and their mean proficiency level in Basque was 6.10 ($SD = 1.75$). (Please note that the potential impact of all this factors is investigated in Section Influence of Linguistic and Nonlinguistic Factors on the Effects).

Group-based pairwise comparisons showed that the two language groups were correctly matched for their age, their overall reading, arithmetical, and attention-related skills (as assessed by their teachers in a Likert-like 1-to-5 scale), and their verbal, nonverbal, and composed IQ (according to the Spanish version of the Kaufman Brief Intelligence Test, K-BIT). Furthermore, the matching was not only done at the general monolingual versus bilingual levels, but also at the individual grade level, as shown in Table 1. None of the t-tests resulted in significance (all ps > .35 at the group level and all ps > .20 at the grade level). A series of strict criteria was followed for the final inclusion of the participants in the experiments. First, none of the participants in any group had any specific deficit, disorder, or special education needs (this was attested by a questionnaire completed by the parents or legal tutors and by the teachers). Second, none of the participants had repeated any academic year. Third, all of the participants had reading, arithmetical, or attention-related skills that were rated with scores equal to or higher than 2 in the 1-to-5 Likert-like scales. Fourth, none of the 504 participants scored below or equal to the 20th centile in the verbal, nonverbal, and composed IQ tests. Hence, given the strict inclusion criteria and the careful matching of the participants, we believe that little doubt could be cast regarding the similarity between test groups in all controlled factors except for their linguistic profile.

Materials

Eight Spanish words were used in the classic verbal Stroop task: The names of the colors green, red, blue, and yellow (''verde,'' ''rojo,'' ''azul,'' and ''amarillo'' in Spanish), and four pairwise-matched words with a similar length, frequency, and syllabic structure that did not correspond to color names (''torno,'' ''sala,'' ''olor,'' and ''uniforme,'' translated as drill or lathe, lounge, smell, and uniform, respectively). These words were then arranged to create the Congruent, Incongruent, and Neutral Word conditions. The Congruent condition (24 trials) was created by presenting each of the color names printed in the color that matched the lexical entry (e.g., the word ''verde'' printed in green ink). In the Congruent condition each color name was presented six times (i.e., 4 color names × 6 presentations = 24 trials). The Incongruent condition (24 trials) was created by presenting each color name printed in a color that did not match the color represented by the lexical entry (e.g., the word ''verde'' printed in red ink). To this end, each color name was presented printed in each of the other colors twice (i.e., 4 words × 3 colors × 2 presentations = 24 trials). The Neutral Word condition (24 trials) was created by presenting the noncolor words in the ink color that corresponded to their pairwise-matched

Table 1. Characteristics of the samples of monolingual and bilingual children tested in Experiments 1 and 2

Grade		Age (in years)		Reading scores (1–5)		Math scores (1–5)		Attention scores (1–5)		Verbal IQ (centiles)		Non-verbal IQ (centiles)		General IQ (centiles)	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Primary school 3rd grade	Monolinguals	8.02	0.35	4.21	0.9	4.31	0.78	4.21	0.84	77.62	17.57	66.17	21.92	70.55	19.62
	Bilinguals	8.05	0.38	4.12	0.89	4.36	0.76	4.19	0.83	76.26	18.79	65.93	22.38	69.83	19.73
	p-value	0.66		0.62		0.77		0.89		0.63		0.93		0.59	
Primary school 4th grade	Monolinguals	9.05	0.38	4.69	0.78	4.74	0.89	4.6	0.94	67.29	19.48	68.07	19.13	65.14	17.5
	Bilinguals	9	0.22	4.71	0.77	4.74	0.86	4.6	0.91	66.57	18.77	67.43	19.52	64.76	19.17
	p-value	0.49		0.86		1		1		0.78		0.8		0.76	
Primary school 5th grade	Monolinguals	10	0.22	4.67	0.75	4.64	0.73	4.43	0.83	58.71	17.59	68.95	17.06	61.1	16.37
	Bilinguals	10.05	0.31	4.62	0.76	4.62	0.82	4.48	0.77	55.98	18.85	69.21	18.94	60.36	17.3
	p-value	0.42		0.74		0.88		0.76		0.32		0.93		0.58	
Primary school 6th grade	Monolinguals	10.98	0.41	4.76	0.82	4.71	0.81	4.71	0.89	60.38	19.77	69.02	18.21	62.36	17.42
	Bilinguals	11	0.44	4.81	0.77	4.6	0.7	4.74	0.99	59.86	20.25	68.52	18.74	61.64	18.75
	p-value	0.8		0.78		0.4		0.9		0.87		0.86		0.6	
High school 1st grade	Monolinguals	12.05	0.31	4.19	0.74	4.14	0.68	4.36	0.96	49.98	17.22	65.1	17.86	53.5	18.42
	Bilinguals	12.07	0.34	4.05	1.01	4.21	0.95	4.38	1.15	50.9	16.65	67.69	18.29	55.95	17.2
	p-value	0.74		0.36		0.63		0.91		0.72		0.32		0.21	
High School 2nd grade	Monolinguals	12.93	0.46	4.5	0.86	4.24	0.93	4.38	0.94	70.07	18.79	72.9	14.45	70.17	16.04
	Bilinguals	13	0.54	4.45	0.89	4.19	1.09	4.38	0.94	68.74	17.04	73.69	15.23	70.1	16.04
	p-value	0.52		0.73		0.78		1		0.55		0.73		0.92	
Total	Monolinguals	10.5	1.73	4.5	0.83	4.46	0.83	4.45	0.91	64.01	20.29	68.37	18.23	63.8	18.38
	Bilinguals	10.53	1.75	4.46	0.89	4.45	0.89	4.46	0.95	63.05	20.11	68.75	18.93	63.77	18.61
	p-value	0.45		0.49		0.85		0.87		0.38		0.72		0.96	

counterparts from the color name set. As in the Congruent condition, each word was presented six times (i.e., 4 words × 6 presentations = 24 trials). Finally, we also included a Control Symbol condition (24 trials) in order to be able to explore potential differences between groups with a minimal influence from reading-related processes (see Results section). To this end, strings of percentage symbols (e.g., "%%%%%") were presented in the four possible ink colors (i.e., 4 colors × 6 presentations = 24 trials). Hence, each participant was presented with a total of 96 experimental trials. The trial presentation order was randomized across participants.

Procedure

The students were tested in their schools and the same technological equipment was used in the data collection across sites (same PCs, same peripherals). The experiment was run using DMDX (Forster & Forster, 2003) and verbal responses were collected through Sennheisser PC151 headsets. Trained research assistants helped in the data collection process and all data was gathered at the schools during teaching hours. Participants were first presented with a recording of the instructions via headphones. They were instructed to name the color of the ink of each of the strings presented on the screen. Next, the experimenters asked the participants whether they had comprehended the instructions, and in those cases in which participants did not fully understand the task requirements, they were again presented with the recording. Following this procedure, we managed to minimize the potential impact of experimenter-driven differences in the recording sessions. After the instructions, participants completed a short familiarization phase that included four trials (one per condition), and received feedback regarding their accuracy in the practice trials. Immediately after this, participants were presented with the 96 experimental trials. Participants first saw a fixation mark that was briefly displayed in the center of the screen for 250 ms and once the fixation mark disappeared, the visual display containing the experimental item was presented until a verbal response was given or for a maximum of 2,500 ms. All the strings were presented in uppercase Courier New font on a black background. The precise RGB-scale values for each of the colors of the ink of the words were as follows: green = 0, 255, 0; blue = 0, 0, 255; red = 255, 0, 0; yellow = 255, 255, 0. The whole experimental session lasted around 8 min.

Results

Individual verbal responses were collected and resulting data were preprocessed and corrected for incorrect voice key triggering with the help of CheckVocal (Protopapas, 2007). Incorrect responses (< 2% of the data) and reaction times below or above 2.5 standard deviations from the mean in each condition for each participant (< 2.5% of the data) were excluded from the latency analysis. The mean latencies for correct responses and error rates are presented in Table 2. Different ANOVAs were conducted in order to explore (1) the classical Stroop effect (Incongruent vs. Congruent trials), (2) the incongruity effect and its interaction with the linguistic profile (Incongruent vs. Neutral Word trials), and (3) the congruency effect and its interaction with the linguistic profile (Congruent vs. Neutral Word trials).[1] Given the strong claims that could be derived from any experiment testing different samples, we decided to meet statistically strict criteria and assess the degree of generalization of the results using the minF' statistic that allows for simultaneous generalization of the results at the participant and item level (which corresponds to the lower bound of the F'; see Clark, 1973).[2]

First, we explored whether there were significant differences between all the conditions across participants (a general Condition effect) and whether the effect of Language or the interaction was significant. To this end, we performed a series of ANOVAs including the factors Condition (four levels: Congruent, Incongruent, Neutral Words, Neutral Symbols) and Language (two levels: Monolinguals, Bilinguals). A significant effect of Condition was found, min$F'(3, 152) = 236.69$, $p < .001$, but the effect of Language was negligible, min$F'(1, 525) = 1.43$, $p > .24$. The interaction between these two factors did not approach significance (min$F' < 1$, $p > .37$). Next, we decided to explore the individual Stroop, incongruity, and congruity effects separately. The separated analysis of the different effects is of crucial importance for the general claims of this study, given that the presence or absence of a bilingual advantage in a given index of the Stroop paradigm (e.g., in the analysis of the congruity effects) does not necessarily need to be accompanied by a similar trend in other indices (e.g., in the analysis of the Stroop or incongruity effects). Hence, in the following sections we present the different contrasts and effects (see also Figure 1), following earlier reports by, among others, Bialystok et al. (2008). Besides, in order to explore in depth potential interactions between the linguistic profile of the participants and each of the

[1] We performed a parallel set of analyses using the Neutral Symbol condition as a baseline. However, given that response times and error rates for the Neutral Symbol condition highly resembled those for the Congruent condition (see Table 2), we decided to maintain the Neutral Words condition as a baseline for the analysis, since it allowed for a correct identification of both incongruity and congruency effects. Nonetheless, we wish to stress that none of the analysis of the incongruity or congruency effects performed using the Neutral Symbols condition as a baseline showed any significant effect of Language or interaction between Language and Condition (reaction times: min$F$'s < 1.15, ps > .29; error rates: min$F$'s < 1.80, ps > .18).

[2] Some of the analyses regarding the net effects reported in the manuscript (see Complementary Analysis Section) do not allow for by-item analysis, and consequently the min$F$' statistic cannot be calculated. In those cases, the by-participant F statistic is reported.

Table 2. Mean reaction times (in ms) and error rates (percentage) in all conditions tested in Experiment 1 for the monolingual and bilingual groups

	Reaction times																
	Conditions										Effects						
	Congruent		Incongruent		Neutral word		Neutral symbol		Total		Stroop		Congruency		Incongruity		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Bilinguals	784	136	977	181	892	167	781	127	855	140	193	105	108	99	−85	86	
Monolinguals	771	137	963	176	870	163	769	129	841	140	191	107	98	93	−93	86	
	Error rates																
	Conditions										Effects						
	Congruent		Incongruent		Neutral word		Neutral symbol		Total		Stroop		Congruency		Incongruity		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	
Bilinguals	0.48	2.26	5.04	7.43	1.82	4.24	0.69	2.34	2.01	3.18	4.56	7.02	1.34	3.78	−3.22	6.47	
Monolinguals	0.89	2.61	4.60	6.41	1.54	3.84	0.83	2.26	1.96	2.81	3.70	6.27	0.64	3.80	−3.06	5.82	

effects tested, we decided to include the factor Language in the analyses. Nonetheless, it should be clearly noted that the negligible effect of Language and the null interaction between Language and Condition in the omnibus ANOVAs would already make this analysis unnecessary. However, we chose to include it to provide the reader with further confirmation of the minimal role that bilingualism plays in this task.

Classic Stroop Effect

We performed a series of ANOVAs including the factors Condition (Incongruent vs. Congruent) and Language (Bilingual vs. Monolingual). In the *reaction time* analysis, we observed a significant effect of Condition, min-$F'(1, 76) = 377.00$, $p < .001$, showing that the incongruent trials were responded to slower than the congruent ones. We did not find a main effect of Language, min-$F'(1, 546) = 1.00$, $p > .31$, and this factor did not interact with Condition (min$F' < 1$, $p > .85$) (see Figure 1). The classic Stroop effects were highly similar for the two Language groups (Cohen's $d = .02$).

The analysis on the *error rates* revealed a main effect of Condition, min$F'(1, 187) = 103.22$, $p < .001$, showing that participants made more errors in the incongruent than in the congruent trials. However, the Language effect was not significant (min$F' < 1$, $p > .96$). The Language by Condition interaction was not significant, min$F'(1, 338) = 1.47$, $p > .22$. Again, the classic Stroop effects were highly similar for the two Language groups (Cohen's $d = .13$).

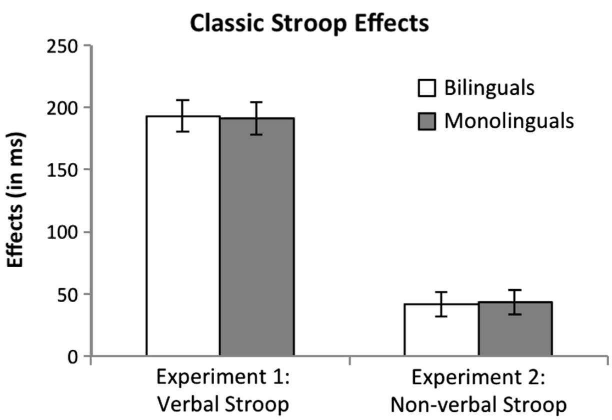

Figure 1. Net reaction time effects (in ms) for the classic Stroop index obtained from the monolingual (gray bars) and bilingual children (white bars) tested in Experiments 1 and 2. Error bars represent the 95% confidence intervals.

Incongruity Effect

A different series of analyses was performed in order to explore the incongruity effect (Incongruent vs. Neutral Word). We decided to use as a baseline condition the Neutral Word condition instead of the Neutral Symbol condition, given that we wanted to avoid different processing biases related to the processing of linguistic versus nonlinguistic material. ANOVAs on the *reaction times* revealed a main effect of Condition, min$F'(1, 70) = 103.95$, $p < .001$, but no effect of Language, min$F'(1, 536) = 1.51$, $p > .21$, nor interaction between Language and Condition (min-$F' < 1$, $p > .34$) (see Figure 2). The incongruity effects were highly similar for the two Language groups (Cohen's $d = .09$).

ANOVAs on the *error rates* revealed a significant main effect of Condition, min$F'(1, 152) = 61.42$, $p < .001$, and no effect of Language or interaction (minF's < 1 and ps $> .43$). The incongruity effects were virtually identical for the two Language groups (Cohen's $d = -.03$).

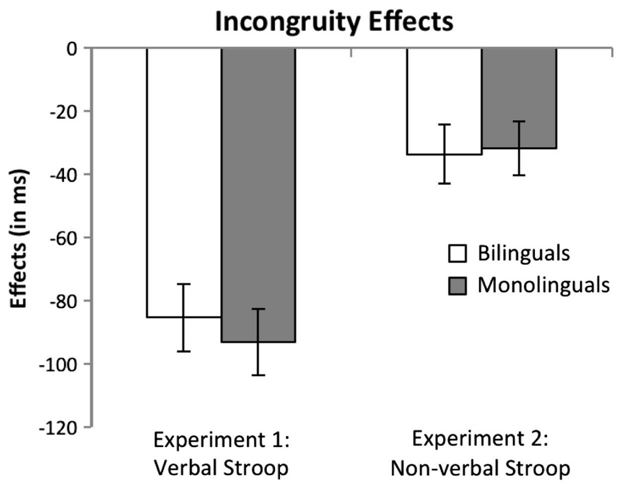

Figure 2. Net reaction time effects (in ms) for the incongruity effects obtained from the monolingual (gray bars) and bilingual children (white bars) tested in Experiments 1 and 2. Error bars represent the 95% confidence intervals.

Congruency Effect

In order to explore the facilitation (i.e., congruency) effect, we performed a two-way ANOVA including the factors Condition (Congruent vs. Neutral Word condition) and Language (bilinguals vs. monolinguals). In the *reaction time* analysis, a significant main effect of Condition was found, min$F'(1, 99) = 189.37$, $p < .001$. However, the main Language effect was not significant, min-$F'(1, 538) = 1.81$, $p > .17$ nor was the interaction between Condition and Language, min$F'(1, 448) = 1.01$, $p > .31$. There was a significant facilitation effect (shorter reaction times for congruent stimuli than for neutral stimuli) but this difference was similar for bilinguals and monolinguals. The congruency effects were similar for the two Language groups (Cohen's $d = .10$).

ANOVAs on the *error data* showed a main effect of Condition, min$F'(1, 176) = 17.68$, $p < .001$, showing that participants made more errors in the neutral condition than in the congruent condition. The main effect of Language was not significant (min$F' < 1$, $p > .82$). The interaction between Language and Condition was not significant either, min-$F'(1, 309) = 2.83$, $p = .09$. The congruency effects were similar for the two Language groups (Cohen's $d = .18$).

Experiment 2: Numerical Stroop Task

Method

Participants

The same participants as in Experiment 1 took part in this experiment.

Materials

Forty-eight visual displays were created. The same number of trials was used for each of the experimental conditions (congruent: 16 trials; incongruent: 16 trials; neutral: 16 trials). In order to create the visual displays in each condition, eight digits were used throughout the experimental session (1, 2, 3, 4, 6, 7, 8, and 9). Each digit was presented the same number of times during the experimental session (12 times). Furthermore, each digit was used the same number of times in each condition (4 times). Participants were presented with visual displays including two digits, one on the left side and another one on the right side. As stated, the visual displays could correspond to one out of three conditions: congruent, incongruent, or neutral. In congruent trials, participants were presented with a pair of two different digits (e.g., 1–6) and the digit representing the largest numerical magnitude (e.g., 6) was presented in a larger font size than the digit representing the smallest numerical magnitude (e.g., 1). Hence, in these trials the numerical and the physical information of the visual inputs coincided. In incongruent trials, the size of the digit representing the smallest numerical magnitude was the largest (e.g., a big 1 and a small 6). This way, the numerical and the physical information provided by the digit pairs was incongruent in these trials. Finally, in neutral trials the same digit was presented at each side of the visual display, but in a different font size (e.g., a big 1 and a small 1). This way, in the absence of any congruent or incongruent numerical information, participants could make their decisions simply based on the physical information. In each type of trial (congruent, incongruent, neutral), half of the visual displays required a right-hand response and the other half required a left-hand response.

Procedure

All the technical equipment and software used in this experiment were identical to those reported in Experiment 1. The instructions were also given via headphones using a previously recorded audio file. Participants were instructed to decide which of the two digits displayed on the screen was larger in size by pressing a key in the keyboard (right digit: L key; left digit: S key). After hearing the instructions, participants completed three practice trials (one per condition) and feedback regarding their accuracy was provided. Immediately after the practice trials, the experimental trials were presented in a random order for each participant. First, a fixation mark was presented in the center of the screen for 300 ms in order to capture participants' attention. Next, the visual display was presented until participants had responded to it or for a maximum of 3,500 ms. All the digits were presented in Courier New black font on a white background. The whole experimental session lasted around 5 min.

Results

Incorrect responses (< 2.5% of the data) and reaction times below or above 2.5 standard deviations from the mean in

Table 3. Mean reaction times (in ms) and error rates (percentage) in all conditions tested in Experiment 2 for the monolingual and bilingual groups

	Reaction times													
	Conditions								Effects					
	Congruent		Incongruent		Neutral		Total		Stroop		Congruency		Incongruity	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Bilinguals	696	206	737	193	703	191	712	188	42	80	8	55	−34	76
Monolinguals	683	194	727	188	695	179	701	176	44	80	12	67	−32	70
	Error rates													
	Conditions								Effects					
	Congruent		Incongruent		Neutral		Total		Stroop		Congruency		Incongruity	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Bilinguals	1.04	2.76	3.82	4.99	1.24	3.16	2.03	2.32	2.78	5.58	0.20	3.94	−2.58	5.68
Monolinguals	0.94	2.50	4.54	7.75	0.89	3.23	2.12	3.08	3.60	8.14	−0.05	3.11	−3.65	8.32

each condition for each participant (< 2.5% of the data) were excluded from the latency analysis. The mean latencies for correct responses and error rates are presented in Table 3. As in Experiment 1, different ANOVAs were conducted in order to explore (1) the classic Stroop effect, (2) the incongruity effect and its interaction with the linguistic profile, and (3) the congruency effect and its interaction with the linguistic profile. The same statistical approach followed in Experiment 1 was used.

As in Experiment 1, we first ran an ANOVA including the factors Condition (3 levels: Congruent, Incongruent, Neutral) and Language (2 levels: Monolinguals, Bilinguals). The Condition effect was significant, min-$F'(2, 89) = 28.68$, $p < .001$, but the Language effect and the interaction between these factors were negligible (min-F's < 1 and *p*s > .51). Next, we explored each of the individual effects. As in Experiment 1, it should be noted that the negligible effect of Language and the null interaction between Language and Condition in the omnibus ANOVAs would discredit the inclusion of the factor Language in the following analyses. Nonetheless, we decided to present those data in order to convince the skeptical reader.

Classic Stroop Effect

ANOVAs on the *reaction times* exploring the factors Condition (Congruent vs. Incongruent) and Language (monolinguals vs. bilinguals) showed a main effect of Condition, min$F'(1, 67) = 48.12$, $p < .001$, but no Language effect or interaction (minF's < 1 and *p*s > .51). Participants took longer to respond to incongruent stimuli than to congruent stimuli (see Figure 1). The Stroop effects were highly similar for the two Language groups (Cohen's $d = -.03$).[3]

The ANOVAs on the *error data* showed a main effect of Condition, min$F'(1, 55) = 28.12$, $p < .001$, and no effect of Language (min$F' < 1$, $p > .44$) or interaction, min-$F'(1, 172) = 1.05$, $p > .30$. Participants made more errors in the incongruent condition than in the congruent condition. The Stroop effects were similar for the two Language groups (Cohen's $d = -.11$).

Incongruity Effect

We compared responses in the incongruent condition to those in the neutral condition. ANOVAs on the *reaction times* showed a significant effect of Condition, min-$F'(1, 53) = 25.61$, $p < .001$. The Language effect and the interaction were not significant (minF's < 1, *p*s > .56). Longer reaction times were found for incongruent trials than for neutral trials (i.e., an incongruity effect; see Figure 2). The incongruity effects were similar for the two Language groups (Cohen's $d = -.08$).

3 We also explored potential interactions between the SNARC effect (i.e., the finding that small numbers elicit faster left responses and that large numbers elicit faster right responses) and the magnitude of the effects in each group. The development of the SNARC effect is somewhat controversial, but most studies suggest that while the effect is attenuated in Grades 1 and 2, it becomes relatively stable already in Grade 3 (Berch, Foley, Hill, & Ryan, 1999; Girelli, Lucangeli, & Butterworth, 2000; Van Galen & Reitsma, 2008). For each test group (monolinguals and bilinguals), we calculated the magnitude of the Stroop effects for congruent and incongruent trials that respected the mental number line (namely, higher values located at the right side; eight trials per condition). Results showed identical effects for the two groups (Monolinguals: Congruent = 680 ms, Incongruent = 736 ms, Stroop effect = 56 ms; Bilinguals: Congruent = 692 ms, Incongruent = 747 ms, Stroop Effect = 55 ms). A parallel procedure was followed for trials in which the location of the digits on the screen did not match the mental number line (i.e., higher values located at the left side; eight trials per condition), and in spite of a clear generalized SNARC effect, again no differences were found between the two groups of participants in the magnitude of the effects (Monolinguals: Congruent = 686 ms, Incongruent = 721 ms, Stroop effect = 35 ms; Bilinguals: Congruent = 700 ms, Incongruent = 732 ms, Stroop Effect = 32 ms).

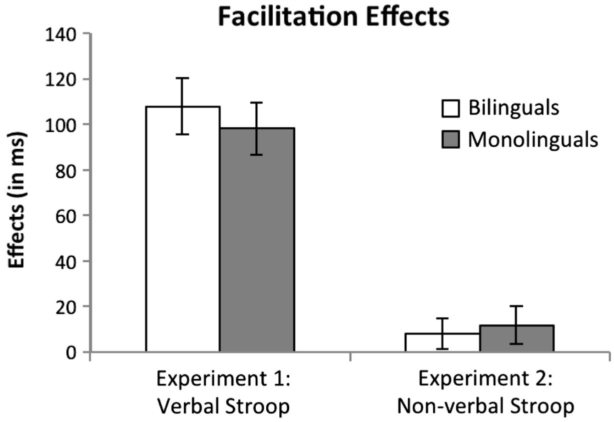

Figure 3. Net reaction time effects (in ms) for the congruency effects obtained from the monolingual (gray bars) and bilingual children (white bars) tested in Experiments 1 and 2. Error bars represent the 95% confidence intervals.

The analysis on the *error data* showed a significant main Condition effect, $\min F'(1, 57) = 26.62$, $p < .001$. The Language effect was not significant ($\min F' < 1$, $p > .65$). The interaction between these two factors was not significant, $\min F'(1, 185) = 1.77$, $p > .18$. The incongruity effects were also similar for the two Language groups (Cohen's $d = .15$).

Congruency Effect

ANOVAs on the *reaction time* data exploring the facilitation or congruency effect (i.e., Congruent vs. Neutral conditions) showed a marginal main effect of Condition, $\min F'(1, 57) = 3.63$, $p = 0.06$, but no effect of Language or interaction ($\min F'$s < 1, ps $> .51$). Congruent trials were responded faster than neutral trials (see Figure 3). The incongruity effects were also similar for the two Language groups (Cohen's $d = -.07$).

The analysis on the error data showed no main effect of Language ($\min F' < 1$, $p > .41$). The Condition effect and the interaction were also negligible ($\min F'$s < 1, ps $> .60$). The congruency effects were virtually identical for the two Language groups (Cohen's $d = .07$).

Complementary Analysis

Considering the importance of the current results for a better understanding of the extent to which inhibitory skills diverge in monolingual and bilingual samples of children, we decided to perform a series of complementary analyses investigating other factors that could be hiding the differential effect that was not found in Experiments 1 and 2. To this end, four analyses are reported. First, we explored whether the lack of bilingual advantage found in the two experiments held constant across all the groups of participants in all the grades tested. Second, we investigated whether the lack of differences extended to the distribution of reaction times across all conditions in the two experiments. Third, we explored whether some factors associated with the linguistic background of the bilinguals (i.e., percentage of exposure, age of acquisition, and proficiency level) could have influenced the results. And fourth, we investigated whether there were overall differences in the general performance in each of the tasks between monolinguals and bilinguals.

Grade-Based Analysis

We analyzed the net effects obtained in Experiments 1 and 2 taking into account the grade the children were when they participated in the experiments (see Tables 4 and 5). In Experiment 1 (classic Stroop task), the *net classic Stroop effect* was calculated, and an analysis was performed including the factors Grade and Language, in order to explore how this effect developed across the six different grades and two language profiles. Analysis of the RT data showed a main effect of Grade, $F(5, 492) = 23.64$, $p < .001$, and no effect of Language or interaction between these factors (Fs < 1 and ps $> .71$). This demonstrates that the Stroop effect was larger in younger than in older groups and that the development of the effect was similar for both bilinguals and monolinguals. Analysis of the error data

Table 4. Classic Stroop effects, congruency effects, and incongruity effects for each monolingual and bilingual age group tested in Experiment 1. Net RT effects are reported in ms, and net error rate effects are reported in percentage (within parentheses)

	Stroop effect				Congruency effect				Incongruity effect			
Grade	Bilinguals		Monolinguals		Bilinguals		Monolinguals		Bilinguals		Monolinguals	
Primary school 3rd grade	275	(9.52)	269	(8.83)	198	(3.47)	174	(2.78)	−77	(−6.05)	−96	(−6.05)
Primary school 4th grade	225	(6.75)	205	(3.97)	113	(2.08)	123	(0.30)	−112	(−4.66)	−82	(−3.67)
Primary school 5th grade	216	(4.66)	213	(3.47)	142	(1.09)	119	(0.10)	−75	(−3.57)	−94	(−3.37)
Primary school 6th grade	153	(2.78)	168	(2.18)	62	(0.60)	70	(0.10)	−91	(−2.18)	−98	(−2.08)
High school 1st grade	145	(1.59)	164	(2.38)	71	(0.30)	62	(0.00)	−74	(−1.29)	−103	(−2.38)
High school 2nd grade	145	(2.08)	128	(1.39)	62	(0.50)	42	(0.60)	−82	(−1.59)	−86	(−0.79)

Table 5. Classic Stroop effects, congruency effects, and incongruity effects for each monolingual and bilingual age group tested in Experiment 2. Net RT effects are reported in ms, and net error rate effects are reported in percentage (within parentheses)

Grade	Stroop effect				Congruency effect				Incongruity effect			
	Bilinguals		Monolinguals		Bilinguals		Monolinguals		Bilinguals		Monolinguals	
Primary school 3rd grade	24	(2.98)	19	(2.53)	−7	(0.45)	3	(−0.45)	−31	(−2.53)	−16	(−2.98)
Primary school 4th grade	46	(1.34)	38	(3.42)	1	(1.19)	27	(0.89)	−45	(−0.15)	−11	(−2.53)
Primary school 5th grade	53	(4.17)	62	(3.72)	11	(0.89)	11	(−0.60)	−42	(−3.27)	−51	(−4.32)
Primary school 6th grade	56	(3.57)	46	(2.98)	13	(0.15)	8	(0.15)	−43	(−3.42)	−38	(−2.83)
High school 1st grade	36	(1.93)	39	(3.13)	16	(−0.45)	4	(0.00)	−20	(−2.38)	−35	(−3.13)
High school 2nd grade	35	(2.68)	58	(5.80)	13	(−1.04)	16	(−0.30)	−22	(−3.72)	−41	(−6.10)

showed a parallel pattern, with a main effect of Grade, $F(5, 492) = 17.67$, $p < .001$, and no effect of Language, $F(1, 492) = 2.447$, $p > .11$, nor interaction ($F < 1$, $p > .59$) (see Table 4). Similarly, an analysis on the *net incongruity effects* across grades and linguistic profiles was performed for the data collected in Experiment 1. For the RTs we did not find any specific modulation of the incongruity effect as a function of Grade ($F < 1$, $p > .88$) or Language, $F(1, 492) = 1.01$, $p > .31$, nor interaction between these factors, $F(5, 492) = 1.20$, $p > .30$. The analysis on the net incongruity effect in the error rates showed a main effect of Grade, $F(5, 492) = 7.59$, $p < .001$, but no significant Language effect ($F < 1$, $p > .75$) or interaction ($F < 1$, $p > .90$). Participants made more errors in the incongruent condition than in the neutral one and this difference was smaller for the older participants than for the younger ones. Finally, we investigated whether the *net congruency effect* varied across age and language groups in Experiment 1. The analysis on the RT effects showed a main effect of Grade, $F(5, 492) = 31.15$, $p < .001$, but no effect of Language, $F(1, 492) = 1.68$, $p > .19$, nor interaction between Grade and Language ($F < 1$, $p > .61$). The difference between the reaction times to the neutral and the congruent stimuli decreased over time. A parallel analysis on the facilitation effect in the error rates showed a main effect of Grade, $F(5, 492) = 7.53$, $p < .001$, revealing that the net congruency effect decreased over time. The Language effect was significant, $F(1, 492) = 4.48$, $p < .04$, showing that the congruency effect was larger for bilinguals than for monolinguals but the two factors did not interact with each other ($F < 1$, $p > .65$).

We also performed an analysis on the *net Stroop effect* obtained in Experiment 2 adding the factor Grade to the design (see Table 5). For RTs, we found a marginal effect of Grade, $F(5, 492) = 2.01$, $p = .075$, and no effect of Language or interaction between Grade and Language (Fs < 1, ps $> .76$). These results moderately suggest that the Stroop effect increased with age but demonstrate that it was similar for monolingual and bilingual children. A parallel analysis on the net Stoop effect found in the error rates did not show any significant effects (Language effect: $F(1, 492) = 1.73$, $p > .18$; Grade effect: $F(5, 492) = 1.02$, $p > .40$; interaction: $F(5, 492) = 1.06$, $p > .38$). Next, we performed an analysis on the *net incongruity effect* obtained in Experiment 2 including the factor Grade. For the net RT effects, we found no effect of Language ($F < 1$, $p > .77$), Grade, $F(5, 492) = 1.22$, $p > .29$, or interaction, $F(5, 492) = 1.64$, $p > .14$. For the net error rate effects, we found a main effect of Grade, $F(5, 492) = 2.37$, $p < .04$, and no effect of Language, $F(1, 492) = 2.85$, $p = .09$, and no interaction between them ($F < 1$ and $p > .73$). The Grade effect demonstrated that the general incongruity effect augmented as a function of age for both language groups. We also performed an analysis on the *net congruency or facilitation effect* adding the factor Grade. For the RT data, we found no effect whatsoever (all Fs < 1 and ps $> .48$). For the net effects in the error rates we found a marginal effect of Grade, $F(5, 492) = 2.16$, $p = .06$, but no effect of Language ($F < 1$, $p > .42$) or interaction, $F(5, 492) = 1.18$, $p > .31$. The magnitude of the net congruency effect decreased with age.

Reaction Time Distributions

As noted by Ratcliff (1979), analyzing reaction time distributions can be appropriate in order to uncover potentially masked effects that are not evident at first sight in analyses performed on the averages of different conditions. This is of special relevance in psycholinguistic studies, given that reaction time distributions do not follow a normal (Gaussian) distribution (see Balota & Spieler, 1999; Luce, 1986). Hence, some effects may be particularly linked to long or short RTs, and consequently analyses on the distributions of the RTs are highly recommended (see Roelofs, Piai, & Rodriguez, 2011).

We computed the RT distributions in each critical condition tested in Experiments 1 and 2. Figure 4 shows these distributions, highlighting that the pattern of RTs was highly similar for bilinguals and monolinguals in all test conditions. In order to statistically confirm this observation, we obtained five representative centiles (namely, the 10th, 30th, 50th or median, 70th and 90th centiles; see Table 6). Next, we performed a series of ANOVAs including the factors Language (monolinguals, bilinguals) and Centile (10th, 30th, 50th, 70th, 90th) on the RT data from each condition in each experiment.

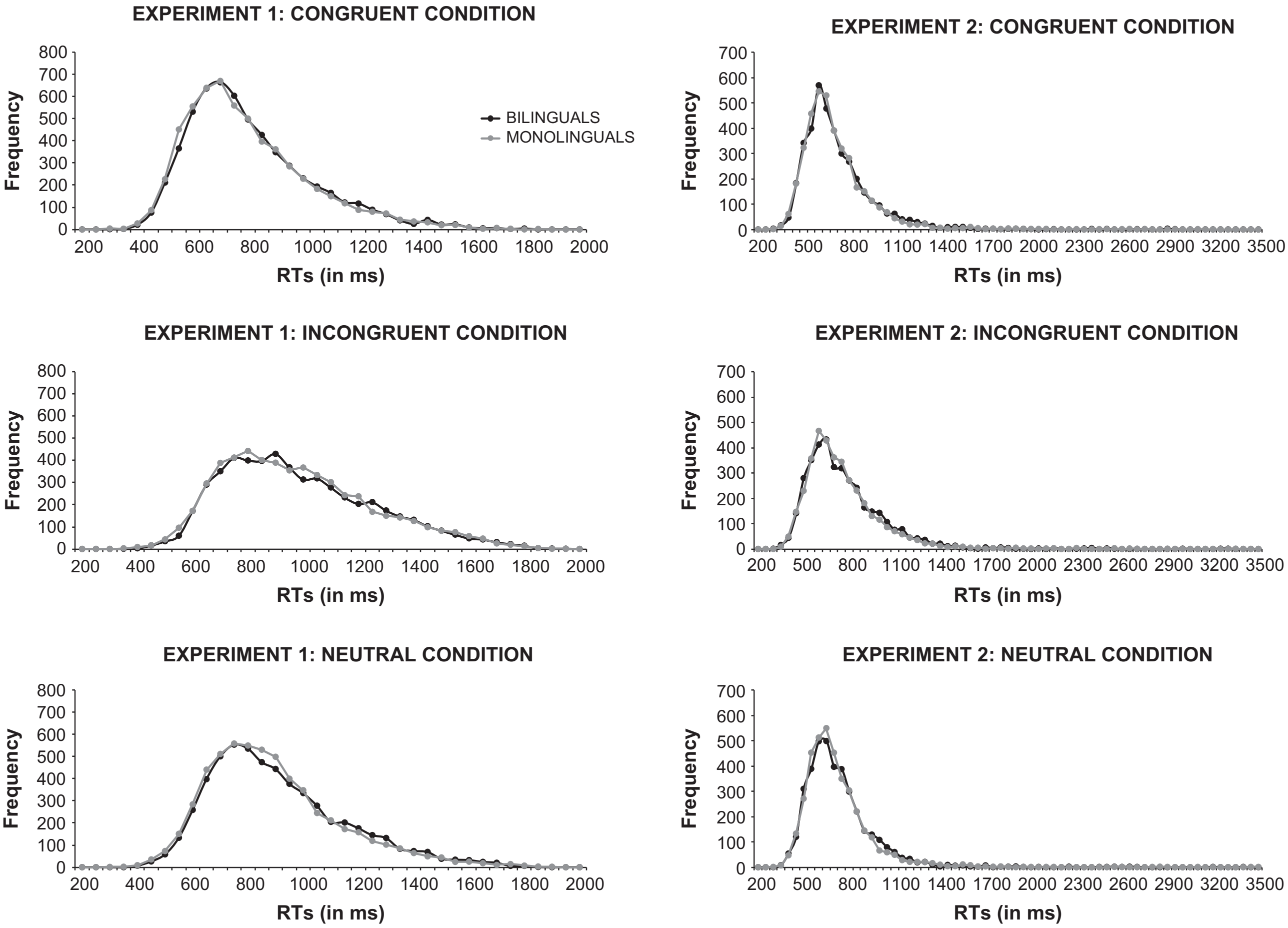

Figure 4. Reaction time distributions for the three critical conditions tested in Experiments 1 and 2. Data from monolinguals are reported in gray and data from bilinguals in black.

Results from the ANOVA on the Congruent condition from Experiment 1 showed an expected effect of Centile, $F(4, 2008) = 2{,}713.10$, $p < .001$, but no effect of Language, $F(1, 502) = 1.09$, $p > .29$, or interaction ($F < 1$, $p > .95$). An ANOVA on the RT data from the Incongruent condition showed a similar pattern, with a significant effect of Centile, $F(4, 2008) = 3{,}345.37$, $p < .001$, and no other effect or interaction (Fs < 1, ps $> .38$]. Finally, an ANOVA on the RT data from the Neutral Word condition in Experiment 1 showed a significant effect of Centile, $F(4, 2008) = 2{,}718.78$, $p < .001$, and no effect of Language, $F(1, 502) = 2.50$, $p > .11$, or interaction, $F(4, 2008) = 1.17$, $p > .31$.

A parallel analysis was performed on the RT data of each condition tested in Experiment 2. The ANOVA on the data from the Congruent condition showed a main effect of Centile, $F(4, 2008) = 870.46$, $p < .001$, and no effect of Language or interaction (Fs < 1, ps $> .43$). Similar results were obtained in the ANOVA on the RT data from the Incongruent condition, with a significant Centile effect, $F(4, 2008) = 1{,}163.96$, $p < .001$) and no effect of Language or interaction (Fs < 1, ps $> .40$). Finally, the ANOVA on the RT data from the Neutral condition showed analogous results, with a significant Centile effect, $F(4, 2008) = 799.42$, $p < .001$, and no Language effect or interaction (Fs < 1, ps $> .62$).

These results confirm that the similarity between the reaction times in all conditions explored in Experiments 1 and 2 followed a markedly similar distribution for the two test groups (monolinguals and bilinguals) and therefore we can safely conclude that the null differences obtained between groups that are reported in previous sections are not masking potential differences condensed in one of the tails of the RT distributions. (Note in this regard that the net effects are computed from the original data in each condition and, given the lack of differences between bilinguals and monolinguals in each condition, no between-group differences are expected in the magnitudes of the effects).

Influence of Linguistic and Non-Linguistic Factors on the Effects

In order to reinforce the view that the effects found in the present study are unlikely to have been modulated by other

Table 6. Distributions of the reaction times in each condition from Experiments 1 and 2 according to the 10th, 30th, 50th (median), 70th, and 90th centiles for monolingual and bilingual participants.

	Experiment 1					Experiment 2				
	Congruent condition					Congruent condition				
	10th Centile	30th Centile	50th Centile	70th Centile	90th Centile	10th Centile	30th Centile	50th Centile	70th Centile	90th Centile
Bilinguals	595	679	758	857	1009	564	623	675	742	853
Monolinguals	585	668	746	840	995	555	612	664	726	839
	Incongruent condition					Incongruent condition				
	10th Centile	30th Centile	50th Centile	70th Centile	90th Centile	10th Centile	30th Centile	50th Centile	70th Centile	90th Centile
Bilinguals	743	858	954	1072	1242	577	642	709	794	936
Monolinguals	731	841	935	1053	1238	576	639	697	777	917
	Neutral (word) condition					Neutral condition				
	10th Centile	30th Centile	50th Centile	70th Centile	90th Centile	10th Centile	30th Centile	50th Centile	70th Centile	90th Centile
Bilinguals	696	793	870	966	1122	575	637	687	749	850
Monolinguals	682	772	847	937	1092	570	626	677	741	845

independent linguistic and non-linguistic factors, we decided to explore whether any of the measures gathered from the bilingual sample significantly modulated the classic Stroop, congruency, and/or incongruity effects in each of the experiments (see Table 7). To this end, we performed a series of regression analyses including as independent factors: (1) the scores reported by the teachers from the 1-to-5 Likert scale regarding reading, mathematical, and attention skills, (2) the verbal, non-verbal, and composed IQ scores in centiles, (3) the percentage of the time exposed to Spanish, (4) the age of acquisition for Spanish and Basque (in years) reported by the parents, and (5) the parents' perception scores (according to a 1-to-10 scale) about their child's proficiency in Spanish and Basque. For (1) and (2) data for the whole set of participants was included, while for (3), (4), and (5) data for 11 participants (representing only 4% of the sample) could not be entered to the analysis, since their parents did not report these scores in the questionnaires. As seen in Table 7, results consistently showed that none of the effects was significantly modulated by any of these factors. Furthermore, none of the regression models were significance (all ps > .24). The explanatory capacity of the independent factors was markedly low (all adjusted R^2s < .015). Hence, we believe that we can safely conclude that the absence of differences between monolinguals and bilinguals does not depend on the specific distribution of the independent measures that were reported, and that the different effects explored are insensitive to those linguistic and non-linguistic factors.

Overall Differences in the Performance

In an attempt to explore between-group differences at other levels different from those represented by the classical markers in these tasks, we also investigated the potential differences in the general performance of bilinguals and monolinguals in Experiments 1 and 2. Some recent evidence has shown that the bilingual advantage could respond to general enhanced monitoring processes that could lead to differences in the overall performance in a given task with high attentional demands (e.g., Costa et al., 2009; see Hilchey & Klein, 2011, for review). According to this idea, between-group differences could be observed in the overall reaction times and error rates, showing that bilinguals perform better in these tasks (faster and/or more accurately) than monolinguals. Therefore, we also explored potential differences in the overall performance of bilinguals and monolinguals in the whole dataset. To this end, the mean reaction times and error rates across the four conditions tested in Experiment 1 and the three conditions tested in Experiment 2 were analyzed. Again, we did not observe any general difference between monolinguals and bilinguals. In Experiment 1, the Language effect in the overall reaction time analysis was negligible, min-$F'(1, 524) = 1.25$, $p > .26$. The Language effect in the overall error data was not significant either (min$F' < 1$, $p > .86$). In Experiment 2, the same negligible Language result was replicated (all minF's < 1 and ps > .51).

Table 7. Results of the series of regression analyses performed over the net classic Stroop effects, the congruency effects, and the incongruity effects for the bilingual sample in each of the experiments

	Stroop effect				Congruency effect				Incongruity effect			
	Experiment 1		Experiment 2		Experiment 1		Experiment 2		Experiment 1		Experiment 2	
ANOVA	$F = 1.27, p = .24$		$F = 1.10, p = .36$		$F = 1.08, p = .38$		$F = 1.23, p = .27$		$F = .76, p = .68$		$F = 1.25, p = .25$	
Adjusted R^2	.013		.005		.004		.010		−.011		.011	
	t-value	*p*-value	*t*-value	*p*-value	*t*-value	*p*-value	*t*-value	*p*-value	*t*-value	*p*-value	*t*-value	*p*-value
Reading skills (1–5)	−0.21	0.84	0.98	0.33	−0.11	0.91	0.41	0.68	0.12	0.90	−0.74	0.46
Mathematical skills (1–5)	−0.25	0.80	1.34	0.18	−0.70	0.49	0.72	0.47	−0.46	0.64	−0.90	0.37
Attention skills (1–5)	−0.15	0.88	−1.23	0.22	−0.73	0.47	−0.62	0.54	−0.63	0.53	0.85	0.40
Verbal IQ (centile)	1.29	0.20	0.69	0.49	−0.15	0.88	−0.97	0.34	−1.74	0.08	−1.43	0.16
Non-verbal IQ (centile)	0.49	0.63	1.27	0.20	−0.72	0.47	0.11	0.91	−1.39	0.17	−1.27	0.21
General IQ (centile)	−0.73	0.47	−1.21	0.23	0.58	0.56	0.41	0.68	1.52	0.13	1.58	0.12
Exposure to Spanish (%)	−1.91	0.06	1.55	0.12	−1.24	0.22	−0.13	0.90	0.96	0.34	−1.73	0.09
AoA Spanish (years)	−0.15	0.88	−0.77	0.44	0.50	0.62	0.44	0.66	0.73	0.47	1.13	0.26
AoA Basque (years)	−1.13	0.26	−0.79	0.43	−1.25	0.21	1.21	0.23	0.00	1.00	1.72	0.09
Proficiency Spanish (1–10)	−0.84	0.40	−0.04	0.97	−0.12	0.91	0.95	0.34	0.89	0.37	0.74	0.46
Proficiency Basque (1–10)	−0.82	0.41	0.23	0.82	−0.43	0.67	0.24	0.81	0.52	0.60	−0.06	0.95

Therefore, in light of these results we can safely conclude that there was no bilingual advantage responding to general monitoring skills.

General Discussion

In recent years several studies have suggested the existence of a so-called bilingual advantage in domain-general cognitive skills (see Bialystok et al., 2012, for summary) with a special emphasis on the positive effects of bilingualism over attention-related skills. While some postulates suggest that the bilingual advantage responds to a bilingual executive processing advantage (e.g., Bialystok et al., 2004; Costa et al., 2009), other hypotheses suggest that the bilingual advantage responds to the enhanced inhibitory control mechanisms of bilinguals (e.g., Bialystok & Craik, 2010; Luk, Anderson, Craik, Grady, & Bialystok, 2010; see Hilchey & Klein, 2011, for review). These positive consequences of multilingualism have been largely explored in adult samples, and they seem to be relatively stable in the elderly (see Luo, Craik, Moreno, & Bialystok, 2013). However, several unsuccessful replications of these studies have suggested that the degree of generalization of these effects to other samples is markedly limited (e.g., Costa et al., 2008; Paap & Greenberg, 2013, regarding inhibitory control). In the current study we aimed at establishing whether bilingual children exhibit enhanced inhibitory skills in a linguistic and a non-linguistic task, due to their daily use of (and switch between) two languages as compared to carefully matched monolingual peers. Overall, the pattern of results obtained does not confidently and reliably allow for such a strong conclusion, given that the evidence favoring a clear-cut difference between monolinguals and bilinguals in inhibitory skills was not found.

We aimed at exploring the influence of the linguistic profile of children on their inhibitory skills in a large sample of monolingual and bilingual children of different ages. Previous evidence regarding children's ability to inhibit conflicting information in a variety of paradigms such as the Stroop task shows a high degree of consistency in the findings, suggesting that monolingual children typically exhibit significant and trustworthy effects based on inhibitory control from early ages (see Montgomery & Koeltzow, 2010, for review). However, to date, there is insufficient data regarding similarities and differences in the magnitude of the classic indices in these paradigms for bilingual children and how they resemble those from monolingual children. Our data show unequivocal interference or incongruity and congruency effects all across the range of ages tested. In both tasks, congruent trials were responded to faster than neutral trials (i.e., a congruency effect), and incongruent trials were responded to slower than neutral ones (i.e., an incongruity effect). Furthermore, overall reaction times decrease as a function of increasing age, so that older participants were faster than younger children. However, the interactions between the magnitudes of these indices (Stroop and incongruity effects) classically associated with inhibitory control and the linguistic profiles of the participants demonstrated that bilingual and monolingual children do not differ in their in inhibitory skills.

The reaction time patterns observed in the classic and numerical versions of the Stroop task (Experiments 1 and 2, respectively) showed a significant generalized Stroop effect (incongruent vs. congruent trials), but critically this effect was highly similar for bilinguals and monolinguals. The negligible difference between monolinguals and bilinguals in the magnitudes of the classic Stroop effect is clearly at odds with preceding studies using a similar index of inhibitory skills (see Qiu, Luo, Wang, Zhang, & Zhang, 2006, for review). Therefore, other fine-grained measures have to be considered in order to explore potential differences between groups, investigating separately congruency and incongruity or interference effects with respect to a neutral condition (see Bialystok et al., 2008; see also Barch et al., 1999; Carter, Mintun, & Cohen, 1995). To this end, we performed a parallel set of analyses exploring the incongruity effect (incongruent vs. neutral), on the one hand, and the congruency effect (congruent vs. neutral), on the other. Comparing each of the critical conditions with a neutral one presents an easy way to disentangle the locus (or loci) of the potential differences between monolinguals and bilinguals. Again, following this strategy, we failed to observe any specific bilingual advantage in the reaction time data: monolingual and bilingual children displayed identical effects. According to recent evidence on the bilingual advantage (see Bialystok et al., 2004; Martin-Rhee & Bialystok, 2008; see Hilchey & Klein, 2011, for review), one would initially have predicted a diminished interference effect in bilingual compared to monolingual children if inhibitory mechanisms and skills were enhanced in bilinguals. In contrast to this prediction, our data suggest that bilinguals performed similarly to their monolingual peers. Thus, the first main conclusion that we can draw from these reaction time data regarding the effects of interference and congruency in the two tasks here tested is that we did not find any clear indication of a bilingual advantage supporting the assumption of enhanced inhibitory skills in this group (see also Paap & Greenberg, 2013, for evidence with young adults). Importantly, when analyzing these data taking into account potential differences between monolingual and bilingual children depending on their age (or grade), as well as other linguistic and non-linguistic factors (see Complementary Analysis Section), we again failed to find any significant effects of the linguistic profile of the participants (nor interaction between Grade and Language), reinforcing the absence of bilingual advantage in inhibitory skills.

When turning to the effects observed in the error rates, the pattern very much resembles that obtained in the reaction time analyses, offering a pattern of data that is inconsistent with the proposal of enhanced cognitive control in bilingual children. The only remarkable differential effect found in the series of analyses performed on the error data was found in the classic Stroop task (Experiment 1). In this task we did not find any supporting evidence for group difference in the Stroop effect or in the incongruity effect when analyzing the error data. However, there seems to be a small differential congruency effect between monolinguals and bilinguals: bilingual children showed a slightly larger congruency effect than their monolingual peers (see also Bialystok et al., 2004), partly due to differences in the baseline. Still, it seems difficult to establish a consistent link between this effect and any sort of enhanced inhibitory skills in bilinguals, given that it corresponds to a difference in congruent trials caused by differences in the baselines (with no differences in the performance of monolinguals and bilinguals in the incongruent trials). However, this small difference in the congruency effect could be potentially accommodated by theories that posit the locus of the bilingual advantage at the general executive functioning level, rather than at the concrete level of inhibitory mechanisms (e.g., de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; Costa et al., 2009). Nonetheless, in the numerical Stroop task (Experiment 2), we did not replicate this difference in the error data associated with the congruency effect. Given the high number of participants tested, the reduced number of errors in these tasks, and the lack of cross-experiment replicability of these effects, the degree of generalization of a potential bilingual advantage based on this error effect is highly limited and great caution is advised in this regard. Furthermore, the series of complementary analyses performed on the data from Experiments 1 and 2 demonstrated that bilinguals performed similarly to monolinguals in these tasks, with no significant differential Language effects in their overall performance (either in the reaction time data or in the error data). This was evidenced by the lack of differences between the two groups in the RT distributions, as well as by the null differences in the analysis taking into account the entire sets of data.

It is worth noting that the cross-task coherence in the magnitude of the effects for both bilingual and monolingual children is surprisingly low. All the correlations between the RT effects (classic Stroop, facilitation, and inhibition effects) were exaggeratedly mild (classic Stroop: $r = .07$; facilitation: $r = -.05$; inhibition: $r = .14$), in spite of the large sample of participants being tested ($N = 504$). We believe that this lack of consistency across indices and tasks confirms the difficulty to generalize arguments based on a bilingual advantage related to inhibitory skills, given the scarce coherence and consistency of apparently similar indices across apparently similar tasks and paradigms (see also Miyake & Friedman, 2012, for a similar argument). Hence, task-specific differences between bilinguals and monolinguals should be taken with extreme caution, given the lack of cross-task convergence that does not grant general claims regarding the implications of bilingualism for inhibitory skills associated with executive processing.

In the current study we demonstrated a relative consistency of the results across two different versions of the Stroop task with markedly different involvement of explicit linguistic knowledge and lexical access. Obviously, the classic Stroop task (Experiment 1) requires a direct selection of a lexical item (i.e., the name of the color to be verbally produced). However, the numerical Stroop task (Experiment 2) does not necessarily involve lexical retrieval, given that the manual response required from the participants relates to a physical property of the items displayed on the screen. It should be considered, however, that the numerical version of the Stroop task may not be completely blind to linguistic representations, given that the linguistic tag associated with each of the Arabic digits presented to the participants could have been activated during the course of the trials (namely, bilingual children may have activated the Spanish and Basque lexical representations of the digit 2, "dos" and "bi," while monolingual children may have exclusively accessed to the lexical form "dos" in the same context). Nonetheless, the way in which this hypothesized distinct degree of lexical activation and dispersion between monolinguals and bilinguals could have influenced the pattern of results does not have a completely transparent and straightforward answer. On the one hand, according to bilingual models of lexical access and/or lexical organization (e.g., the Bilingual Interactive Activation model and its revisions, or the Revised Hierarchical Model; see Grainger & Dijkstra, 1992; Kroll, Van Hell, Tokowicz, & Green, 2010; van Heuven, Dijkstra & Grainger, 1998), the degree of lexical dispersion would be higher in bilinguals than in monolinguals. Whether or not this increased lexical dispersion in bilinguals could have led to faster or slower responses is a matter of debate. Still, it should be clearly stated that, as reported in the analysis exploring the overall differences in the performance, we did not find any significant difference between monolinguals and bilinguals in the response latencies. Furthermore, exclusively considering the neutral condition of the numerical Stroop task (that is not influenced by either congruency or incongruency effects), the RTs were highly similar across groups (703 ms for bilinguals and 695 for monolinguals), and so were the error rates (1.24% and 0.89%, respectively; see Table 3). Hence, taking these observations into account together with the high consistency between the findings reported in Experiments 1 and 2, we believe that these data speak for a relatively stable pattern of results across two tasks with clearly different linguistic demands.

These results may seem surprising at first sight, considering preceding evidence supporting the bilingual advantage in tasks in which inhibitory control mechanisms are directly implicated (see, among others, Bialystok, 2009; Bialystok et al., 2008; Bialystok & Martin, 2004; Craik & Bialystok, 2006). Nevertheless, as already stated earlier, there are other studies showing that this advantage is not uncontroversial, and that differential effects in executive control may largely depend on a number of factors that still remain unclear. Paap and Greenberg (2013) presented evidence obtained from a variety of tasks mainly tapping into inhibitory skills. They did not observe any evidence suggesting a bilingual advantage in a Simon task, in a Flanker task, and in an Antisaccade task. They claimed that there are several external socio-demographic variables that are usually uncontrolled that may have favored the apparent bilingual advantages that had been previously found (but see de Abreu et al., 2012, and Bialystok et al., 2012, for counterarguments), as well as other statistical issues that could have been behind those differential effects. Among other issues, Paap and Greenberg referred to the typically small-sized groups of participants explored as a potential reason guiding the between-group differences. We believe that this issue is correctly dealt with in the present study, given that the sample size was clearly large (252 monolinguals and 252 monolinguals), representing the largest sample tested so far in this regard. Besides, the matching of the groups was done taking into account the age, the scores received from the children's teachers according to their skills in reading, mathematics and attention, and the scores obtained in a general IQ test and the sample selection (see Methods section). We also tried to eliminate any potential influence of socioeconomic status by restricting the inclusion of participants to those who lived in (and were originally from) the same country. Hence, to our eyes, the only relevant and evident difference between groups corresponded to their linguistic profile: while children who were immersed in a bilingual (academic) context formed one of the groups, the other group was made up of children from a purely monolingual context.

As mentioned above, these results may not be totally surprising, considering recent research reporting no differences between monolinguals and bilinguals in inhibitory control, but they are clearly at odds with several earlier studies showing enhanced inhibitory skills in bilinguals. Nevertheless, this study probably is, to our knowledge, the most exhaustive, extensive, and well-controlled bilingual study that has been conducted so far to test and compare monolingual and bilingual children's inhibitory skills. Certainly, we wish to be cautious with regard to the conclusions to be drawn from this study. The present data demonstrate that in these two Stroop tasks the bilingual advantage is not found when children are tested but these data do not necessarily imply that differences stemming from participants' linguistic background cannot emerge later in life (e.g., with adult samples). This is especially critical if one considers the evidence showing that the bilingual advantage is much more evident in old adulthood than in young adulthood (e.g., Bialystok et al., 2007, 2008; but see Linck, Hoshino, & Kroll, 2008). In fact, as Bialystok et al. (2012) point out, differences between monolinguals and bilinguals are more reliable in older samples than in younger ones ("the evidence for a bilingual advantage in younger adults is more sporadic than in other age groups," p. 243). This issue has recently been highlighted by Hilchey and Klein (2011) in their meta-analysis of several studies testing interference effects and multilingualism. Their analysis revealed that the number of studies showing significant differences between monolingual and bilingual children or young adults is certainly reduced, leading the authors to conclude that "the absence of a bilingual advantage in these age groups is simply inconsistent with the proposal that bilingualism has a general positive effect on inhibitory control processes" (p. 629).

Similarly, we would want to stress that the lack of bilingual advantage in these two tasks does not discredit all possible cognitive differences that can be found between monolingual and bilingual children or infants. For instance, a remarkable number of studies have consistently shown clear-cut differences between pre-verbal children from bilingual and monolingual contexts in language-mediated visual discrimination of faces (e.g., Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012; Weikum et al., 2007). The current study does not provide evidence at this regard, but rather focuses on the differences between bilingual and monolingual children in a relatively specific aspect of executive control, demonstrating that the so-called bilingual advantage in inhibitory mechanisms as measured by the two versions of the Stroop paradigm is untrustworthy and unstable. This study adds to a progressively increasing number of studies reporting negligible support for a bilingual advantage in inhibitory control in different groups of bilinguals, and pretends to partially oppose to the well-recognized bias against publishing null effects that is observed in the field of experimental psychology (see Francis, 2012). Further large-scale investigations testing a much larger age range and involving different situations are clearly needed in order to build a clear picture of the influence of multilingualism in domain-unspecific cognitive abilities.

Acknowledgments

This research was partially supported by Grants CSD2008-00048, PSI2010-15133, PSI2011-23340, and PSI2012-32123 from the Spanish Government, ERC-AdG-295362 from the European Research Council, and PI2012-74 from the Basque Government. We wish to thank all the children and their families for kindly collaborating in this project and all the different schools for providing us with infrastructure. We are also extremely thankful to the research assistants who helped us in the data collection, to Margaret Gillon Dowens for her helpful comments, and to Matthew Hilchey and Kenneth Paap for their valuable suggestions on earlier drafts of this manuscript. We are particularly grateful to Maria Teresa Bajo for her priceless comments and contributions.

References

de Abreu, P. M. J. E., Cruz-Santos, A., Tourinho, C. J., Martin, R., & Bialystok, E. (2012). Bilingualism enriches the poor enhanced cognitive control in low-income minority children. *Psychological Science, 23*, 1364–1371.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*, 32–55.

Barch, D. M., Carter, C. S., Perlstein, W., Baird, J., Cohen, J. D., & Schooler, N. (1999). Increased Stroop facilitation effects in schizophrenia are not due to increased automatic spreading activation. *Schizophrenia Research, 39*, 51–64.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology, 74*, 286–308.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology/ Revue canadienne de psychologie expérimentale, 60*, 68–79.

Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*, 3–11.

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and Linguistic Processing in the Bilingual Mind. *Current Directions in Psychological Science, 19*, 19–23.

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

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.

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences, 16*, 240–250.

Bialystok, E., Craik, F. I. M., & Ruocco, A. C. (2006). Dual-modality monitoring in a classification task: The effects of bilingualism and ageing. *Quarterly Journal of Experimental Psychology, 59*, 1968–1983.

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.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7*, 325–339.

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9*, 103–119.

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

Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H215O PET study of Stroop task performance. *NeuroImage, 2*, 264–272.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Colzato, L. S., Bajo, M. T., Van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., & Hommel, B. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 302–312.

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

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*, 59–86.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences, 10*, 131–138.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*, 340–347.

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

Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin and Review, 19*, 975–991.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology, 76*, 104–122.

Grainger, J., & Dijkstra, A. (1992). On the representation and use of language information in bilinguals. In R. J. Haris (Ed.), *Cognitive processing in bilinguals* (pp. 207–220). Amsterdam, The Netherland: Elsevier.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81.
Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition, 13*, 315–325.
Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*, 625–658.
Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica, 127*, 277–288.
Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. *Memory & Cognition, 15*, 531–543.
Kadosh, R. C., Gevers, W., & Notebaert, W. (2011). Sequential analysis of the numerical Stroop effect reveals response suppression. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*, 1243–1249.
Kousaie, S., & Phillips, N. A. (2012). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research, 1446*, 71–90.
Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*, 416–430.
Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13*, 373–381.
Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *The Mental Lexicon, 3*, 349–374.
Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (1st ed.). New York, NY: Oxford University Press.
Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. *Brain and Cognition, 74*, 347–357.
Luo, L., Craik, F. I. M., Moreno, S., & Bialystok, E. (2013). Bilingualism interacts with domain in a working memory task: Evidence from aging. *Psychology and Aging, 28*, 28–34.
Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*, 81–93.
Montgomery, D. E., & Koeltzow, T. E. (2010). A review of the day-night task: The Stroop paradigm and interference control in young children. *Developmental Review, 30*, 308–330.
Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science, 10*, 719–726.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Current Directions in Psychological Science, 21*, 8–14.
Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232–258.
Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*, 859–862.
Qiu, J., Luo, Y., Wang, Q., Zhang, F., & Zhang, Q. (2006). Brain mechanism of Stroop interference effect in Chinese characters. *Brain Research, 1072*, 186–193.
Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446–461.
Roelofs, A., Piai, V., & Rodriguez, G. G. (2011). Attentional inhibition in bilingual naming performance: Evidence from delta-plot analyses. *Frontiers in Psychology, 2*, 184.
Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*, 994–999.
Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*, 300–304.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Experimental Psychology, 18*, 643–662.
Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*, 12530–12535.
Van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology, 101*, 99–113.
van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*, 458–483.
Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*, 1159–1159.

Received June 12, 2013
Revision received August 6, 2013
Accepted August 19, 2013
Published online November 12, 2013

Jon Andoni Duñabeitia

Basque Center on Cognition
Brain and Language (BCBL)
Paseo Mikeletegi 69
2nd floor
20009 Donostia
Spain
Tel. +34943309300
Fax +34943309052
E-mail j.dunabeitia@bcbl.eu