Does bilingualism shape inhibitory control in the elderly?

Eneko Antón[a], Yuriem Fernández García[a], Manuel Carreiras[a, b, c] and Jon Andoni Duñabeitia[a].

[a] BCBL. Basque Center on Cognition, Brain and Language; Donostia, Spain.

[b] Ikerbasque, Basque Foundation for Science; Bilbao, Spain.

[c] Euskal Herriko Unibertsitatea – Universidad del País Vasco; Bilbao, Spain.

Contact information:

Eneko Antón

Basque Center on Cognition, Brain and Language (BCBL)

Paseo Mikeletegi 69, 2nd floor

20009 Donostia, SPAIN

e.anton@bcbl.eu

Telephone: +34943309300

Abstract

Bilingualism has been argued to benefit executive functioning. However, recent research suggests that this advantage may stem from uncontrolled factors or incorrectly matched samples. In this study we test the effects of bilingualism on elderly lifelong bilinguals whose cognitive abilities are in decline, thus making any benefits more salient. Firstly we compare 24 bilinguals and 24 carefully matched monolinguals on verbal and the numerical Stroop tasks, obtaining no differences in monitoring or inhibitory measures. Secondly we explore the modulations that the proficiency in the L2 might cause to executive control functions, as measured by the same tasks, by testing 70 elderly bilinguals who vary in their L2 mastery from very low to perfectly fluent. Results show no modulation in any of the indices due to L2 proficiency. These results add to the growing body of evidence showing that the bilingual advantages might indeed be due to other factors rather than bilingualism.

Keywords: Bilingualism; bilingual advantage; lifelong bilingualism; executive functioning.

Does bilingualism shape inhibitory control in the elderly?

The impact of bilingualism in human cognition has become a popular topic in the past several years and the debate about the so-called bilingual advantage is driving the attention of many researchers in the field. The hypothesis of this apparent advantage of bilinguals over monolinguals in some domain-general non-linguistic skills is grounded on the assumption that bilingualism provides speakers with an additional cognitive training: when it comes to facing communicative situations with different speakers and different demands, bilinguals have to choose between the two languages that they have available in their repertoire to make their communication efficient and to adequate it to the requirements of the current context. Given that the languages that a bilingual speaker speaks are active always (e.g., Lagrou, Hartsuiker, & Duyck, 2013; Midgley, Holcomb, van Heuven, & Grainger, 2008; Thierry, & Wu, 2007), the language that is not meant to be used in a given situation has to be inhibited to avoid any cross-linguistic contamination that could harm effective communication. In contrast, monolinguals do not face these situations, so no language choice is needed and no language inhibition is required. Based on this difference between bilinguals and monolinguals, the hypothesis of the bilingual advantage postulates that bilinguals have increased experience with their executive functions, which is a cognitive construct formed by inhibition (i.e., the ability to suppress dominant responses), shifting (i.e., the ability to switch between tasks), and monitoring (i.e., the ability to update the information in the working memory; see Miyake et al., 2000; Miyake & Friedman, 2012). In other words, given that bilinguals are constantly monitoring their linguistic environment and inhibiting the non-relevant language, the so-called hypothesis of the bilingual advantage suggests that they have enhanced executive functions as compared to monolinguals (e.g., Bialystok, Craik, Green, & Gollan, 2009). In fact, evidence in favor of this advantage has been provided in a variety of studies using non-linguistic executive control tasks, mainly (but not only) focusing on two relevant measurements: inhibitory and monitoring skills (e.g., Bialystok, Craik, Klein & Viswanathan, 2004; Bialystok, Craik & Luk, 2008, 2012; Hernández, Costa, Fuentes, Vivas, & Sebastián-Gallés, 2010; Kroll & Bialystok, 2013).

On the one hand, some researchers have argued that, considering that bilinguals need to reliably inhibit the non-target language, this capacity to avoid intrusion of inappropriate pieces of information would constitute the enhanced skill. Consequently, the advantage should be easily captured by tasks measuring inhibitory skills, such as the classic flanker task (Eriksen & Eriksen, 1974), the Stroop task (Stroop, 1935) or the Simon task (Simon & Rudell, 1967). All these tasks share some features: there are incongruent and congruent trials. Incongruent trials present irrelevant and distracting information and support a different or opposing response from the one expected.

This irrelevant information must either be ignored or the unwanted response needs to be inhibited. Congruent trials are those in which the irrelevant information favors the same response as the one that the relevant information does. Crucially, the difference between the average responses to these two kinds of conditions (i.e., the conflict effect) is found to be smaller in bilinguals (i.e. better inhibitory skills) in some of these tasks (Bialystok, Craik, Klein, & Viswanathan, 2004; Costa, Hernández, & Sebastián-Gallés, 2008; Bialystok, Craik, & Luk, 2008). These differences have been interpreted as a better ability of bilinguals to deal with irrelevant and confusing information and to suppress it, as they do in their daily life when they suppress then non-target language.

On the other hand, the purported bilingual advantage has been also grounded on bilinguals' enhanced monitoring skills (Costa et al. 2009), which stems from the constant need of bilinguals to keep track of the communicative context and monitoring the language needed for the communicative situation in question. These demands would enhance the bilinguals' skill to efficiently tackle taxing contexts in which different stimuli (some relevant, some irrelevant) are present and intermixed, similarly to the everyday situations in which they have to pay attention to different stimuli (the linguistic representations), and select only the relevant ones. In the last few years some researchers have closely related monitoring abilities with global reaction times (RTs) in tasks involving different kinds of congruency conditions, and thus a bilingual advantage in monitoring has been defended empirically by research showing that bilinguals are overall faster than monolinguals when the context of the task is high demanding and that, crucially, this advantage vanishes if the context is low demanding (Bialystok, 2006; Costa et al., 2009; Costa & Sebastián-Gallés, 2014; but see Paap, Johnson, & Sawi, 2014 for a discussion on the impurity of the use of global RTs as a measure of monitoring).

In contrast to the accounts that favor a genuine enhancement of cognitive control as a consequence of bilingualism, a strong and increasing line of research suggests that the so-called bilingual advantage might be due to hidden demographic factors that tend to be differently distributed among the bilingual and monolingual populations under study (e.g., immigrant status, educational level or socio-economic status, among many others; see Paap, Johnson, & Sawi, 2015a), and that these factors, and not bilingualism *per se*, may be responsible for the observed differences. This counter-argument, which constitutes one of the building blocks of opponents to the so-called bilingual advantage, started with the pioneering work by Morton and Harper (2007). Being aware of the crucial role played by many demographic factors in the development and mastery of executive functioning (among many other cognitive skills), they pointed out that preceding research comparing groups with different linguistic contexts had completely neglected the role of these factors. Some years later, other researchers also brought this same idea to the debate, and the importance of possibly confounding socio-demographic factors started to be acknowledged (see Hilchey & Klein, 2011). Good examples of this are provided by socioeconomic status (SES), which has been often correlated with a better performance in executive functioning tasks, as well as by the immigrant status and other ethnicity-related factors, which are known to affect the quality and speed of the performance on executive function tasks (e.g., Mezzacappa, 2004; Noble, Norman, & Farah, 2005). Bearing this in mind, one can easily observe how some of these factors have been overlooked in many studies reporting a bilingual advantage. For instance, Bialystok and Martin (2004) compared Canadian monolingual and bilingual children without measuring SES. Bialystok and Shapero (2005) compared Canadian monolinguals with immigrant bilinguals coming from different linguistic and ethnic background without measuring SES. In this same line, Bialystok, Craik and Luk (2008) mainly included immigrants in their sample of bilinguals (20 out of 24 individuals). More recently, Engel de Abreu and colleagues (Engel de Abreu, Cruz-Santos, Tourinho, Martin, and Bialystok, 2012) matched their monolinguals and bilinguals for SES, but they took each sample from a different country (Portugal and Luxemburg, respectively). All these studies yielded a bilingual advantage, which the authors unambiguously attributed to bilingualism instead of considering the potential effects of the abovementioned factors. Along these lines, in some extensive reviews, Paap and colleagues (see Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015a, 2015b) identified several other deficiencies that may have contribute to the appearance of a bilingual advantage in preceding studies, including not only Type I errors due to inadequately matched groups or small sample sizes, but also uncontrolled external factors or task-dependent effects.

Crucially, when the mentioned confounding factors are controlled for and participants are carefully matched, any sign for a bilingual advantage, either in inhibition (reduced conflict effect) or in monitoring (overall reduced RTs) tends to vanish (see Paap & Greenberg, 2013). In this line, Morton and Harper (2007) tried to replicate the findings obtained by Bialystok, Craik, Klein and Viswanathan (2004) using the Simon task in children, but as opposed to Bialystok et al., they matched both groups on SES, immigrant status and ethnicity. They found no bilingual advantage in this carefully-controlled experimental setting. Following this same rationale, null results (no bilingual advantage) have been replicated in the last several years in studies in which participant groups were carefully matched for the factors mentioned above and in which the number of participants tested was relatively high (see Antón et al., 2014; Duñabeitia et al., 2014; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015a, 2015b; Gathercole et al., 2014).

While methodological concerns might capture the heterogeneity of the results presented so far in studies comparing monolingual and bilingual young adults, an additional critical issue related to the development of certain cognitive skills of interest has been recently raised. Research in this field has mainly focused on studying populations formed by young adults, which are in normal terms at the peak of their domain-general cognitive abilities (around 20-40 years of age, see Hartshorne & Germine, 2015). This fact makes them likely to show a "ceiling effect" moment with respect to their domain-general cognitive abilities, which can mask or hide the presence of a potential difference between bilinguals and monolinguals, and thus the bilingualism-driven differences might arguably be hard to capture.

Therefore, if any relevant cognitive differences are to arise as a consequence of bilingualism, they would be expected to be most salient when the cognitive skills are not at their maximum, but rather when they are still developing (childhood) or already declining (elderly). This is precisely what has been found by Bialystok and her colleagues (Bialystok, Martin, & Viswanathan, 2005): they administrated the Simon task to 5 year-old children, young adults (20 years of age), middle-aged adults (30-59 years of age) and older adults (60-80 years), and found that the bilingual advantage was present in children and middle-aged and older adult groups, but absent in the young adult participants. However, the general picture does not seem to be consistent in these samples either, and the evidence for a bilingual advantage in the childhood and in the elderly has been also questioned recently. While some studies show an advantage for bilingual children as compared to their monolingual peers (e.g., Martin-Rhee & Bialystok, 2008; Yang, Yang, & Lust, 2011), recent findings suggest that, using a carefully matched and large sample, the advantage disappears in children samples, as it is the case with young adults too (Antón et al. 2014; Duñabeitia et al., 2014; Gathercole et al. 2014).

The scenario with bilingual seniors, whose cognitive abilities are already declining due to normal cognitive deterioration caused by age, is also unclear and foggy. There are a relatively reduced number of studies providing seemingly consistent evidence favoring a behavioral advantage in tasks measuring different forms of cognitive control and executive functions in elderly bilinguals. For example, Bialystok, Craik and Luk (2008) showed stronger differences between bilinguals and monolinguals in older samples as compared to young adults in both the Simon and the verbal Stroop tasks (see also Bialystok, Craik, Klein, & Viswanathan, 2004; Gold, Kim, Johnson, Kryscio, & Smith, 2013), but lately other researchers have suggested that those pieces of evidence are not entirely reliable and replicable. For instance, Kirk and colleagues (Kirk, Fiala, Scott-Brown, & Kempe, 2014) tried to replicate the findings obtained by Bialystok et al. (2004) demonstrating a reduction of the Simon effect in older bilinguals, and found neither signs of a bilingual advantage in inhibition nor a difference in global reaction times (i.e., no advantage in monitoring). This inconsistent pattern is also observable in other studies testing elderly participants, where no group differences appear (see Kousaie & Phillips, 2012; see also de Bruin, Bak & Della Sala, 2015, for a study comparing monolinguals and active and passive bilinguals in the Simon Arrow task with no evidence for any bilingual advantage).

Furthermore, defenders of the bilingual advantage have also argued in favor of a different kind of advantage in the elderly by considering bilingualism a form of protection against dementia (i.e., the neuro-protective value of bilingualism; see Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016, for a recent review). For example, Bialystok and colleagues (Bialystok, Craik, & Freedman, 2007) analyzed records of patients that underwent a process of being diagnosed and treated for dementia, and found that bilinguals manifested the incipient symptoms around four years later than monolinguals, with no change in the later rate of progression or course of the illness. However, these same results have not been fully replicated in other samples, and an increasing number of authors deny that the symptomatology of neurodegenerative diseases is delayed in bilinguals due to their seemingly greater cognitive reserve (e.g., Costa & Sebastian-Gallés, 2014; Paap, Johnson & Sawi, 2015a; 2015b). As pointed out by Chertkow and colleagues (Chertkow et al., 2010), 90% of the bilingual patients investigated by Bialystok et al. (2007) were immigrants, and in a replication of these results they found critical interactions between immigrant status and bilingualism. In an attempt to further clarify the scenario, Gollan and her colleagues (Gollan, Salmon, Montoya, & Galasko, 2011) defended that the neuro-protective role of bilingualism is indeed real, but only in people with low educational level. They found that higher degrees of bilingualism were associated with a delay in the diagnosis, but only in bilinguals with a low educational level. As a potential solution for the conundrum presented by the unclear debate of the role of bilingualism in patients with dementia, a handful of studies have opted for a longitudinal approach, testing cohorts starting from a baseline stage in which no signs of dementia are evident. Most of the studies using this approach have shown no consistent delay in the onset of the symptoms caused by bilingualism (Crane et al., 2009; Lawton, Gasquoine, & Weimer, 2015; Sanders, Hall, Katz, & Lipton, 2012; Yeung, St. John, Menec, & Tyas, 2014; Zahodne, Schofield, Farrell, Stern, & Manly, 2014).

This contradictory evidence draws an unclear picture of the potentially beneficial effects that lifelong bilingualism may have in elderly bilinguals in both executive functions and neurodegenerative processes. According to the general view, it seems reasonable to think that if the use of two languages yields any boosting or enhancement of domain-general cognitive abilities, this should be easily captured in elderly lifelong bilinguals, since they have been actively exposed to two languages for decades, consequently training their inhibitory and monitoring skills for much longer than bilingual children or young adults. However, the evidence in this regard is far from clear and additional studies are required to better understand whether or not bilingualism yields beneficial effects for cognitive reserve in the elderly. The current study was designed to shed light on this precise issue.

It has been suggested that instead of focusing on comparisons between bilinguals and monolinguals (i.e., taking bilingualism as an "all or nothing" factor), bilingualism should be better treated as a continuous rather than dichotomous factor. For example, Singh and Mishra (2013) tested high and low proficient Hindi-English bilinguals on a modified saccadic arrow Stroop task and found a group effect, indicating that high proficient bilinguals were faster and had reduced conflict indices, thus showing that L2 proficiency can apparently modulate monitoring and conflict resolution skills. However, some other researchers had also considered this hypothesis and found no differences in inhibitory control, monitoring or switching based on participants' L2 proficiency (see, for example, Paap, Johnson, & Sawi, 2014). If the effect of the proficiency of the L2 found by Singh and Mishra (2013) is robust and consistent, one should be able to easily replicate a similar finding in groups of bilinguals immersed in a society in which the two languages they speak are present and official by exploring how varying degrees of L2 proficiency modulate differently classic markers of inhibitory skills. This was precisely the aim of our second set of experiments.

Here we present evidence from verbal & non-verbal Stroop tasks aimed at exploring if elderly bilingual speakers differ from their monolingual peers in their inhibitory skills and/or in their monitoring abilities; and if among bilinguals there is a modulation based on their L2 proficiency. Considering that many studies have consistently reported that bilinguals suffer a disadvantage in production of spoken responses (Ivanova & Costa, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005) we collected data from both the verbal Stroop task, i.e., the classic color-naming version of the Stroop paradigm (Stroop, 1935) as well as from a less linguistically-charged version of the same paradigm, the numerical Stroop task (Besner & Coltheart, 1979; Kaufmann et al., 2005; Santens & Verguts, 2011). These two paradigms were implemented along the lines that Duñabeitia et al. (2014) followed in their study testing the reliability of the bilingual advantage during childhood, where they found reliable differences between conditions (i.e., the Stroop, congruency and incongruity indices or effects), which were *not* modulated by the language background of the participants. In the current study, in Experiment 1A, we compared monolingual and carefully matched non-immigrant highly proficient lifelong bilingual seniors in the verbal Stroop task, while in the Experiment 1B we compared the same set of participants in the numerical Stroop task. In addition to this, and considering that it has been suggested that bilingualism should be treated as a continuous rather than dichotomous factor (e.g., Singh & Mishra, 2013), in Experiment 2 we explored the impact of the "degree of bilingualism", here measured as non-immigrant bilinguals' general proficiency in their second language, in executive control functions. A wide range of bilingual seniors with varying degrees of second language (L2) proficiency were tested in the verbal Stroop (Experiment 2A) and the numerical Stroop tasks (Experiment 2B), to see if the classic markers of inhibition (as well as of monitoring) were modulated by participants' L2 knowledge and mastery.

Experiments 1A and 1B: Effect of lifelong bilingualism in executive control.

In this first set of experiments, the hypothesis of the bilingual advantage in inhibitory skills was tested by comparing elderly bilinguals' and monolinguals' performance in a verbal (Experiment 1A) and in a numerical version (Experiment 1B) of the Stroop paradigm. Both groups of participants were carefully matched in the potentially confounding factors discussed above (cf. Paap & Greenberg, 2013), including general intelligence, socioeconomic status, immigration and ethnicity, so that the only relevant difference between both groups was their linguistic profile. Additionally, bearing in mind that some researchers have circumscribed the presence of a bilingual advantage to lowly educated samples (see Gollan et al., 2011), most of the participants that were recruited had a relatively low educational level (see below for details). If the previously reported instances of bilingual advantages in similar tasks were not the result of the confounding factors that we controlled for, both groups of participants would show significant differences either in the magnitude of the Stroop effects (i.e., bilinguals should show a reduced Stroop effects as a reflection of their enhanced inhibitory skills) or in the overall reaction times (with bilinguals performing overall faster due to their better monitoring abilities).

Method

Participants

48 seniors (28 females) were recruited in the Basque Country (mean age 69.06, SD=4.56; age range = 61-78). All of them reported to have normal or corrected-to-normal vision and none of them had any history of chronic neuropsychological disorders. Every participant signed a written consent form approved by the Ethics and Research Committees of the Basque Center on Cognition, Brain and Language (BCBL) establish.

Half of the participants (14 female; mean age= 68.75, SD=4.42) were Spanish monolinguals; and the other half (14 female, mean age of 69.38, SD=4.58) were Basque-Spanish bilinguals who use both languages everyday and rate themselves as highly proficient in comprehension and production in Basque (average score of 8.04 (SD=0.95) where 1 is really poor level and 10 is perfectly fluent) as well as in Spanish (average score of 8.67 (SD=1.17) over 10). Apart from selfrated proficiency, which has been extensively used in the literature and reported to accurately account for the actual proficiency (see, among many others, Clark, 1981; Heilenman, 1990; LeBlanc and Painchaud, 1985), a native speaker of both languages interviewed the participants to certify that bilinguals could efficiently speak both languages fluently and that monolinguals were not able to communicate in Basque. All the participants acquired their languages before the age of 12. Bilinguals and monolinguals didn't differ in any demographic factor (all *p*s>.5), including the age at which they quit formal schooling, the IQ percentile based on the scores they obtained in an abridged version of the Kaufman Brief Intelligence Test, K-BIT (Kaufman & Kaufman, 1990), and the scores in the Spanish version of the Mini Mental State Examination (MMSE, see Lobo et al., 1979). Considering that some researchers have suggested that the level of education can modulate the presence of a bilingual advantage (see Gollan et al., 2011), we recruited participants from all the educational strata but mostly from the lowest one: 9 bilinguals and 10 monolinguals only finished primary school, 2 bilinguals and 3 monolinguals completed middle school, 3 bilinguals and 5 monolinguals had a professional training, 6 bilinguals and 4 monolinguals completed high school and 4 bilinguals and 2 monolinguals had completed a university degree. To avoid any cultural difference, they were recruited in the same city (Donostia-San Sebastián) and were non-immigrants. Furthermore, based on self-ratings, they didn't differ significantly in general, speaking and comprehension abilities in Spanish (all *p*s >.6, see Table A.1 for detailed information), which was the language in which they were spoken to and tested in during the whole process.

-- Table A.1 around here –

Materials

Experiment 1A. For the *verbal Stroop* task (which was a variation of the classic Stroop task; Stroop, 1935), the names of the colors red, blue and yellow ("rojo", "azul" and "amarillo" in Spanish) and three pairwise-matched non-color words of a similar length, frequency and syllabic structure ("ropa", "avión" and "apellido", translated as clothes, plane and surname, respectively) were arranged to create the Congruent (a color name printed in the same color that it indicates, e.g., the word "rojo" in red), Incongruent (a color name printed in a different color, e.g., the word "rojo" in blue) and Neutral Word (words that weren't color names printed in the three colors, e.g., the word "ropa" in red) conditions, and a Neutral Symbol (a string of dollar signs printed in the three colors, e.g., "$$$$$" in red) condition was added to have a condition unaffected by language. Each condition consisted in 24 trials, presenting each color the same amount of times, paired equally to every word in each condition. Each participant was presented with a total of 96 experimental trials, and the trial presentation order was randomized across participants.

Experiment 1B. For the *numerical Stroop* task, 48 stimuli were created using eight digits (1, 2, 3, 4, 6, 7, 8 and 9). Each digit was presented the same number of times in each condition (4 times) and in total (12 times). Each trial consisted in two digits (e.g., 1-6), one on the left side and another one on the right side of the screen. Three conditions were created: 16 congruent trials (the larger number in magnitude was also the bigger in size, e.g., small 1-big 6), 16 incongruent trials (the larger number in magnitude was the smaller in size, e.g., big 1-small 6) and 16 neutral trials (two same numbers different in size, e.g., big 1-small 1). In all the conditions “left” and “right” responses were equally distributed.

Procedures

All the participants were tested in the BCBL facilities in Donostia-San Sebastián. The experiment was run using DMDX (Forster & Forster, 2003) following the same procedure reported by Duñabeitia and colleagues (Duñabeitia et al., 2014).

Experiment 1A. For the *verbal Stroop task*, verbal responses were collected through Sennheisser PC151 headsets. Research assistants indicated to the participants that they had to name the color of the ink in which the word on the screen was presented, as quickly and as accurately as possible. They completed a short training phase that consisted of four trials, one per condition. Immediately after this, the experiment began. They first saw a fixation mark for 250ms and then the target word appeared on the screen for 3000 ms. All the strings were presented in uppercase Courier New font on a black background, and the colors were set in the RGB-scale values as follows: blue=0,0,255; red=255,0,0; yellow=255,255,0.

Experiment 1B. For the *numerical Stroop*, all the technical equipment and software used in this experiment were identical to that reported for the verbal Stroop task. In this occasion, participants were instructed to indicate with the keyboard which of the digits in each pair displayed in the screen was bigger in size, by pressing “L” to indicate “right” side and “S” to indicate “left” side of the screen. After instructions, they 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 300ms in order to capture participants’ attention. Next, the visual display was presented until participants had responded to it or for a maximum of 3500ms. All the digits were presented in Courier New black font on a white background, each digit of each pair on one side of the screen. The whole experimental session lasted around 5 minutes. The order of the tasks was counterbalanced.

Results

Experiment 1A. For the v*erbal Stroop task,* first latencies were analyzed. Reaction times above and below 2.5 standard deviations from each participant's mean in each condition (< 3.2% of the data) were excluded from the analysis. With the remaining data a general ANOVA was run including the factors Condition (Congruent, Incongruent, Neutral Words and Neutral Symbols) and Group (Monolinguals and Bilinguals). In this general 4x2 ANOVA, we found a significant effect of Condition [$F(3, 136)= 66.88$, $p<.01$] but neither the main effect of Group nor the interaction (all *F*s<1). All the descriptive values of conditions across groups are detailed in Table A.2.

- *Table A.2 around here –*

To explore all the possible venues in which differences could emerge, we also computed 2x2 ANOVAS to analyze the potential differences between the groups in the *classic Stroop effect* (i.e., Congruent vs. Incongruent conditions), *incongruity effect* (Neutral Word vs. Incongruent conditions) and *congruency effect* (Neutral Word vs. Congruent conditions)[1]. We further tested these differences with the Bayesian Null Hypothesis Testing (Rouder et al., 2009; Wetzels et al., 2011), comparing bilinguals to monolinguals for every index (i.e., comparing the indices of *Stroop, congruency* and *incongruity* of both groups) testing the H_0 of no differences against the H_1 that says that bilinguals should have smaller indices than monolinguals (i.e., a test of the so-called bilingual advantage) using Bayesian t-tests.

Stroop effect was significant [$F(1, 46)= 114.95$, $p<.01$] but there was no effect of Group nor an interaction between them (all *F*s<1), which was also supported by the Bayesian t-test (Bayesian t-test of the index between groups: $BF_{01}> 5.55$). The same pattern was obtained for the *incongruity*

[1] In order to explore the possibility that the bilingual advantage may be circumscribed to low-educated bilinguals (cf. Gollan et al., 2011), a reanalysis of the data was done with the subset of participants with the lowest educational level. To this end, the 9 bilinguals and 10 monolinguals that had only completed primary school were selected, and the ANOVAs on the RT data failed to show a significant effect of Group, nor an interaction between the two main factors [*F*s<1]. The analysis of each index reinforced this result, showing a sizeable *Stroop effect* [$F(1,17)=55.40$, $p<.01$], a marginal *incongruity effect* [$F(1,17)= 4.25$, $p<.06$], and a significant *congruency effect* [$F(1,17)=46.12$, $p<.01$], which did not interact with the factor Group [all *F*s<1.55 and *p*s>.23]. The analysis of the error rates mimicked these same results, with all the main effects of Group or interactions with this factor resulting negligible [all *F*s<1].

effect, with significant main effect [$F(1, 46)= 46.42$, $p<.01$) with no effect of Group neither an interaction (all *F*s<1) and no differences between groups when the *incongruity effect* was compared across groups using Bayesian approach (Bayesian t-test of the index between groups: $BF_{01}>4.1$). *Congruency effect* was also significant [$F(1, 46)= 30.29$, $p<.01$] but neither the main effect of Group nor the interaction was significant (all *Fs*<1), and the null hypothesis was supported by the Bayesian t-test (Bayesian t-test of the index between groups: $BF_{01}>5.04$).

The error rate analysis also showed a quite similar pattern, first with a general 4x2 ANOVA and then comparing each index separately by using both 2x2 ANOVAs and Bayesian Null Hypothesis Testing between groups. A general 4x2 ANOVA showed a significant effect of Condition [$F(3, 138)= 8.15$, $p<.01$] and no main effect of Group nor an interaction (*F*s<1). In the analysis of the *Stroop index*, we observed a main Condition effect [$F(1, 46)=7.27$, $p<.02$] but neither the main effect of Group nor the interaction was significant (*F*s<1), indicating no differences in this index between groups (Bayesian t-test of the index between groups: $BF_{01}>3.48$). When the *congruency* effect was explored, we observed no effect of Condition, Group or an interaction (all *F*s<1), and Bayesian analysis comparing the index across group indicated that the null was slightly more likely than the alternative hypothesis (Bayesian t-test of the index between groups: $BF_{01}>1.42$). Finally, the analysis of the *incongruity effect* showed a main effect of Condition [$F(1, 46)=87.53$, $p<.01$] but no main effect of Group or interaction (*F*s<1), and a Bayes factor analysis that favored the null hypothesis over the alternative (Bayesian t-test of the index between groups: $BF_{01}>3.92$).

Experiment 1B. The same rationale as in the previous task was followed in the analysis of the *numerical Stroop* task results. In the latencies, responses above and below 2.5 standard deviations from each participant's mean in each condition (< 2.9% of the data) were excluded from the analysis. After trimming, a general 3 x 2 ANOVA was run with the factors Condition (Congruent, Incongruent and Neutral) and Language (Monolinguals and Bilinguals). Analysis showed only a main effect of Condition [$F(2, 92)= 35.07$, $p<.01$, all other *F*s<1]. See Table A.2 for detailed information for each condition.

As in the previous task, we also explored the *classic Stroop effect* (i.e., Congruent vs. Incongruent conditions), the *incongruity effect* (Neutral vs. Incongruent conditions) and the *congruency effect* (Neutral vs. Congruent conditions)[2] using both ANOVAs and Bayesian Null Hypothesis Testing to compare the indices between groups. *Stroop* and *incongruity* effects showed the same pattern, with strong Condition effects (all *p*s<.01) and no other main effect or interaction (all *F*s<1), while Bayesian analysis also showed that the Null Hypothesis of no differences was more likely than the alternative one of smaller effects for bilinguals (BF_{01}>3.6 and BF_{01}>5.72 for the Bayesian t-tests of the *Stroop* and *incongruity* effects, respectively). In the analysis of the *congruency* effect we also found a main effect of Condition [*F*(1, 46)=7.25, *p*<.02], no effect of Group (*F*<1) and no interaction [*F*(1, 46)=2.91, *p*>.1], but the results coming from the Bayesian Null Hypothesis testing didn't allow us to reach any conclusion (Bayesian t-test of the index between groups: BF_{01}>0.58).

In a similar vein, the general 3x2 ANOVA on the error rates indicated that there was a strong effect of Condition [*F*(2, 92)=7.23, *p*<.01] but no Group effect nor an interaction (*F*s<1). When the indices were explored individually, we observed that the *Stroop effect* was significant [*F*(1, 46)= 8.00, *p*<.01] but no main effect of Group nor an interaction was found (all *Fs*<1), coherent with the results from the Bayesian t-test (Bayesian t-test of the index between groups: BF_{01}>3.48). *Incongruity* effect analysis showed a significant effect of Condition [*F*(1, 46)= 6.67, *p*<.02) and no effect of Group or interaction (all *F*s<1), while Bayesian analysis supported the null hypothesis (Bayesian t-test of the index between groups: BF_{01}>3.92). When the *congruency effect* was analyzed, neither the main effect of Condition [*F*(1, 46)= 1.00, *p*>.32], nor the main effect of Group [*F*(1, 46)= 1.00, *p*>.32] neither an interaction between them was found [*F*(1, 46)= 1.87, *p*>.18]; and Bayesian factor analysis showed that the null hypothesis was slightly more likely than the alternative one (Bayesian t-test of the index between groups: BF_{01}>1.42).

Interim conclusion

Results of the two tasks in this first experiment clearly show no differences between the monolingual and the bilingual group in the critical measures of both inhibitory (*Stroop effect*) and monitoring skills (overall reaction times). Importantly, the tasks worked as expected, replicating

[2] As in Experiment 1A, a reanalysis was run on those participants who only completed the obligatory primary school (9 bilinguals and 10 monolinguals). The ANOVAs on the RT data did not show any significant effect of Group or an interaction between the two main factors [*F*s<1.22 and *p*s>.3]. The analysis of each index reinforced this result, showing a significant *Stroop effect* [*F*(1,17)=10.50, *p*<.01], and a significant *incongruity effect* [*F*(1,17)= 9.42, *p*<.01]. The *congruency effect* was not significant [*F*<1], and none of the interactions of these indices with the factor Group resulted significant [all *F*s<1.06, all *p*s>.31]. The analysis of the error rates mimicked these same results, with all the main effects of Group or interactions with this factor resulting negligible [all *F*s<2.45, and *p*s>.14].

previous findings on the main indices in both reaction times and error rates with significant and strong *Stroop effects*, mainly due to an *incongruity effect*. Importantly, when carefully matched monolinguals and bilinguals were compared using the conservative Bayesian approach, results clearly favor the null hypothesis, indicating that the data is much more likely to be explained by the hypothesis of "no differences" than by any other alternative model. Furthermore, the potential impact of the educational level on the emergence of the so-called bilingual advantage was also considered. While most of the participants tested were not highly educated, this was not the case for all of them, and considering that the differential effects could be stronger in low-educated samples (see Gollan et al., 2011), we ran an additional set of analyses including only the participants with the lowest education levels and parallel results were obtained.

However, instead of a clear cut dichotomous distinction between monolinguals and bilinguals, it might be the case that L2 proficiency modulates the effect. To investigate this potential modulation, we conducted a second experiment using the same tasks with a different sample: a set of bilingual participants who differ in L2 proficiency, ranking from a very limited knowledge to perfectly fluent and balanced bilinguals.

Experiment 2A and 2B: Effect of L2 proficiency in executive control.

Method

Participants

70 Basque-Spanish bilingual seniors (45 females; mean age of 69.36, SD=4.40; age range = 61-81) were recruited from the same city from the Basque Country (Donostia-San Sebastián) and were non-immigrants. All of them rated themselves as highly proficient in Spanish (average rating in a 1-to-10 scale was 8.72; SD=1.08) while they showed as a group great variability in their Basque General Proficiency, ranging from 1 (very poor level) to 10 (perfectly fluent), which was also confirmed in the personal interviews guided by bilingual native speakers from the research center. All of them had acquired their two languages before the age of 12 (see Table A.3 for detailed information about participants' profiles). None of them reported history of chronic neuropsychological disorder(s) and all of them had normal or corrected-to-normal vision. All participants signed the written informed consent form approved by the Ethics and Research Committees of the Basque Center on Cognition, Brain and Language (BCBL), and they completed a cognitive screening that included, as in Experiment 1, the Spanish version of the MMSE (Lobo et al., 1979) and an abridged version of the Kaufman Brief Intelligence Test, K-BIT (Kaufman & Kaufman, 1990). Also, these participants were recruited from every different educational strata: out

of the 70 participants, 20 had only completed primary school, 7 had completed middle school, 12 had a professional training, 12 had completed high school, and 19 got a university degree.

-- *Table A.3 around here* --

Materials

Materials used for Experiments 2A and 2B were the exact same ones used in the Experiment 1A and 1B respectively.

Procedures

The procedure followed for the Experiments 2A and 2Bwas the same that we followed in the Experiment 1A and 1B respectively.

Results

In this experiment we performed the same analysis used in Experiment 1, but instead of a between-subject factor separating bilinguals from monolinguals in two groups, we considered Basque General Proficiency as a covariate. This rating scale varied from 1 (very poor level of Basque) to 10 (very fluent) among our 70 bilingual speakers (see Figures B.1 and B.2).

Experiment 2A. In the v*erbal Stroop task,* reaction times above and below 2.5 standard deviations from each participant's mean in each condition (< 2.9% of the data) were excluded from the analysis.

With the remaining latencies, a four way ANOVA (Condition: Congruent, Incongruent, Neutral Symbol and Neutral Words) was conducted, showing a main effect of Condition [F(3, 207)=168.69, p<.01]. Paired t-tests showed that *Stroop* [t(69)=17.37, p<.01], *incongruity* [t(69)=-10.33, p<.01] and *congruency* [t(69)=7.96, p<.01] effects were significant (see Table A.4 for descriptive values).

After this general analysis of our bilingual participants, a four way ANCOVA was run including the factor Condition (Congruent, Incongruent, Neutral Words and Neutral Symbols) and using the Basque General Proficiency as covariate[3]. A significant effect of Condition [F(3, 204)=

[3] Considering previous findings that show a relation between the intelligence scores and the different executive functioning tasks such as the Stroop task (Adelman, 2002; Arffa, 2007), we also ran a four way ANCOVA including both Basque General Proficiency and IQ percentile values (obtained from the K-BIT) as covariates. Results show that there is a main effect of Condition [F(3, 201)=21.02, p<.01], IQ [F(1, 67)=10.55, p<.01] and an interaction between them [F(3, 201)= 7.29,

18.29, *p*<.01] was found, but the effect of Basque General Proficiency was not significant [*F*(1, 68)=2.11, p>.15], and it didn't interact with Condition (*F*<1).

We also explored each index separately by two-way ANCOVAS to see if there was any modulation of the covariate in the effect. The *classic Stroop effect* (i.e., Congruent vs. Incongruent conditions) was significant [*F*(1, 68)=29.49, *p*<.01] but the main effect of Basque General Proficiency was not [*F*(1, 68)= 1.05, p>.31], and Basque General Proficiency did not modulate the main effect of Condition (*F*<1). The *incongruity effect* (Neutral Word vs. Incongruent Conditions) followed the same pattern, with main effect of Condition [*F*(1, 68)=16.74, *p*<.01], no effect of Basque General Proficiency [*F*(1, 68)=2.06, p>.16] and no modulation of the Basque General Proficiency in the main effect of Condition (*F*<1). Finally, the *congruency effect* (Neutral Word vs. Congruent conditions) showed a marginal effect [*F*(1, 68)=2.96, *p*<.1], with no effect of Basque General Proficiency [*F*(1, 68)=1.96, *p*>.17] nor an interaction between the two main effects [*F*(1, 68)=1.20, *p*>.28].

In the error rate analysis, the general four-way ANCOVA showed that none of the effects or interactions were significant [all *F*s<2 and *p*s>.17].

Considering that preceding studies have proposed the existence of a close relationship between the educational level of the participants and their performance in Stroop-like tasks (see Moering, Schinka, Mortimer, & Graves, 2004), and taking into account that the so-called bilingual advantage has been claimed to depend on this factor (cf. Gollan et al., 2011), we decided to run an additional analysis in order to shed light on this issue. A four-way ANCOVA was run including the factor Condition (Congruent, Incongruent, Neutral Words and Neutral Symbols) and using the Basque General Proficiency and Education (i.e., the age at which participants quit formal education) as covariates. Results showed a significant main effect of Condition [*F*(3,201)=9.06, *p*<.01] and a marginal effect of Education [*F*(1,67)=3.17, *p*=.08], with no effect of Basque General Proficiency [*F*(1,67)=1.85, *p*>.18], nor an interaction between Condition and any of the covariates [all *F*s<1]. When each index was explored independently in the corresponding set of two-way ANCOVAs, the

---

p<.01]. When looking at each index, analysis of the *Stroop effect* revealed a main effect of Condition [*F*(1, 67)=48.86, *p*<.01], IQ [*F*(1, 67)=4.86, *p*<.01] and an interaction between them [*F*(1, 67)=15.25, *p*<.01] indicating that the *Stroop effect* was smaller for higher IQ values (r= -0.44, *p*<.01, n= 70). We observe a similar pattern in the *congruency effect*, with a significant Condition [*F*(1, 67)=19.08, *p*<.01] and IQ [*F*(1, 67)= 7.21, *p*<.01] effects as well as an interaction between them [*F*(1, 67)= 18.16, *p*<.01] indicating that the *congruency effect* was smaller for higher IQ scores (r= -.47, *p*<.01, n= 70). Finally, *incongruity effect* analysis revealed a main effect of Condition [*F*(1, 67)= 7.27, *p*<.01] and IQ [*F*(1, 67)= 15.19, *p*<.01] but no other effect was significant (all *F*s<1). Basque General Proficiency was not significant (all *F*s<1, all *p*s>.33) for any of the indices, nor did it interact with condition.

Stroop effect (Congruent vs. Incongruent) resulted significant [*F*(1,67)=15.45, *p*<.01], but it was not modulated by Basque General Proficiency or Education [*Fs*<1]. The main effects of Basque General Proficiency and Education were not significant either [*Fs*(1,67)<2.6, *p*s>.11]. The *incongruity effect* (Incongruent vs. Neutral Word) was significant [*F*(1,67)=8.71, *p*>.01], but it was not modulated by Basque General Proficiency or Education [all *F*s<1]. The main effect of Education was marginal [*F*(1,67)=3.02, *p*=.09], but the effect of Basque General Proficiency was negligible [*F*(1,67)=1.81, *p*>.18]. Finally, the *congruency effect* (Congruent vs. Neutral Word) was not significant [*F*(1,67)=1.54, *p*>.22], and it was not modulated by Basque General Proficiency [*F*(1,67)=1.14, *p*<.3] or Education [*F*<1]. The main effects of Basque General Proficiency and Education were not significant [*F*(1,67)=1.73, *p*>.19 and *F*(1,67)=2.54, *p*>.12, respectively]. The four-way ANCOVA run on the accuracy data showed that none of the effects or interactions was significant (all *F*s<1.8 and *p*s>.2). Altogether, the marginal main effects of Education that emerged in the general ANCOVA showed that overall reaction times tended to be shorter for people with higher educational level, but critically, this analysis demonstrated that Education did not modulate any of the indices of interest.

To further check for any possible modulation of the indices as a function of the relevant demographic data collected, different correlation analyses were run between the factors of interest and the indices obtained (both for RTs and error rates). Crucially for the purposes of this study, Basque General Proficiency did not correlate with the *Stroop* (r=.07, p>.58), *incongruity* (r=-.003, p>.98) or *congruency* (r=.13, p>.28) indices in the RTs (see Figure B.1). The error rate analysis showed the same pattern, and neither the *Stroop* (r=.04, p>.74), nor the *congruency* (r=-.05, p>.66) and the *incongruity* (r=.06, p>.64) indices were correlated with the general proficiency that participants had in Basque (Footnote 3).

- *Figure B.1 around here* -

Experiment 2B. Reaction times above and below 2.5 standard deviations from each participant's mean in each condition (< 3.2% of the data) of the *numerical Stroop task* were excluded from the analysis. After trimming, a three way ANCOVA was run including the factor

Condition (Congruent, Incongruent and Neutral) and Basque General Proficiency as a covariate[4] (see Table A.4 for descriptives). The main effect of Condition was significant [*F*(2, 136)= 7.39, *p*<.01], but Basque General Proficiency was not, and it did not interact with Condition either (all *Fs*<1). Two-way ANCOVAS to account for each effect and its modulation by Basque General Proficiency, if any, showed that both the S*troop* [*F*(1, 68)=12.54, *p*>.01] and the *congruency* [*F*(1, 68)=7.81, *p*<.01] indices were significant but Basque General Proficiency and its interaction with Condition were not (all *F*s<1). The *incongruity* effect was not statistically significant [*F*(1, 68)=2.93, *p*<.1], and Basque General Proficiency was not and it did not interact with the Condition effect (*Fs*<1).

- *Table A.4 around here* -

The general three-way ANCOVA run in error rates showed that neither Condition [*F*(2, 136)=.17, *p*>.84] nor Basque General Proficiency nor their interaction were significant (*Fs*<1).

As in the Experiment 1, the potential effect of Education (i.e., the age at which participants quit formal education) was investigated in a three-way ANCOVA including Basque General Proficiency and Education as covariates. The main effect of Condition [*F*(2,134)=5.40, *p*<.01] and Education [*F*(1,64)=4.61, *p*<.04] resulted significant. Crucially, the effect of Condition did not interact with any of the other factors [all *F*s<1.5, all *p*s>.22], and the effect of Basque General Proficiency was not significant [*F*<1]. The analysis of the *Stroop effect* showed significant effects of Condition [*F*(1,67)=7.78, *p*<.01] and Education [*F*(1,67)=5.22, *p*<.03], with no other relevant effects or interactions [all *F*s<1]. Similarly, the analysis of the *incongruity effect* showed a main effect of Condition [*F*(1,67)= 4.86, *p*<.04] and of Education [*F*(1,67)=4.52, *p*<.04], with no main effect of Basque General Proficiency [*F*<1] or interaction between Condition and Basque General

[4] As in Experiment 2A, we ran a general ANCOVA including both Basque General Proficiency and IQ percentile values as covariates. Condition [*F*(2, 134)=12.37, *p*<.01], IQ [*F*(1, 67)=14.54, *p*<.01] and the interaction between them [*F*(2, 134)= 6.08, *p*<.01] were significant. Exploring each index, the analysis of the *Stroop effect* revealed a main effect of Condition [*F*(1, 67)=18.43, *p*<.01], IQ [*F*(1, 67)=16.50, *p*<.01] and an interaction [*F*(1, 67)=5.78, *p*<.02], showing that the *Stroop effect* decreased as IQ values increased (r= -0.28, *p*<.02, n= 70). In the *congruency effect* we found a significant IQ effect [*F*(1, 67)= 10.89, *p*<.01], with no other significant main effects or interactions (all *F*s<2, all *p*s>.21). The *incongruity effect* analysis revealed a main effect of Condition [*F*(1, 67)= 10.71, *p*<.01] and IQ [*F*(1, 67)= 15.53, *p*<.01], as well as an interaction [*F*(1, 67)=8.04, *p*<.01], indicating reduced *incongruity effects* for higher IQ values (r= -.34, *p*<.01, n=70). Basque General Proficiency was not significant in any of the analyses, nor did it interact with Condition (all *F*s<1 and *p*s>.53).

Proficiency [*F*<1], neither between Condition and Education [$F(1,67)=2.16$, $p<.15$]. The analysis of the *congruency effect* showed no main effect or interactions [all *F*s<1.5 and *p*s>.24], except for a marginal effect of Education [$F(1,67)=3.8$, $p<.06$]. The main effects of Education that consistently emerged in these analyses showed that participants who quit formal education later were the ones associated with faster reaction times in the different conditions. The three-way ANCOVA on the accuracy data showed that none of the main effects or interactions were significant [all *F*s<1].

Additionally, the possible relationship between the demographic variables of interest and the indices measured in this Numerical Stroop task was explored in a correlation analysis. Crucially for the hypothesis explored in this study, we observed that the *Stroop* (r=.03, p>.8), *congruency* (r=-.1, p>.41) and *incongruity* (r=.08, p>.5) indices were not correlated with the General Basque Proficiency in the RT analysis (see Figure B.2). A similar pattern was observed for the error rates, with none of the indices being correlated with the general proficiency that participants had in Basque (*Stroop*: r=.13, p>.30; *congruency:* r=.10, p>.40; *incongruity*: r=.09, p>.44; Footnote 4).

- *Figure B.2 around here* -

Interim Conclusion

In this experiment we obtained significant and strong Stroop effects in the latency analysis of both tasks, mainly due to the incongruity effects. However, when the impact of Basque General Proficiency in the different indices was analyzed, and even when the variance provided by factors such as IQ or Educational Level was controlled for, the ANCOVAs showed no significant effect of participants' knowledge of a second language neither a modulation of the main indices based on the this knowledge, as measured by their Basque General Proficiency.

General Discussion

In the current study, we analyzed the effects derived from lifelong bilingualism in domain-general cognitive abilities related to inhibitory control and monitoring skills in samples of elderly bilinguals and monolinguals. In Experiments 1A and 1B, we compared monolingual Spanish-speaking old adults and elderly Spanish-Basque bilinguals who have been immersed in a bilingual society and who have used their two languages during the vast majority of their lives on a daily basis. All of them were non-immigrants coming from the same city and did not differ in any of the demographic factors nor in the linguistic skills in Spanish (the language that both groups shared and the language in which they were tested). These participants were presumably in a declining process of their cognitive abilities due to normal aging, although their cognitive functioning was at normal levels according to the scores obtained in the Mini Mental State Examination (MMSE, see Folstein,

Folstein & McHugh, 1975; i.e., all participants scored above 26 in Experiment 1 (Median = 29.5), and above 24 in Experiment 2 (Median = 29). These bilinguals were tentatively selected as a good test case to explore any enhancing effect that bilingualism may have in the inhibitory control and monitoring skills, because they have been exposed to lifelong bilingualism and they are not at their peak cognitive abilities. The results unambiguously demonstrated a complete absence of differences between lifelong bilingual seniors and their monolingual peers either in monitoring abilities (which would have been reflected in overall faster reaction times) or in inhibitory skills (which would have been shown by reduced conflict or Stroop effects). Importantly, when the same hypotheses were tested only in the subsets of seniors with the lowest educational levels following Gollan et al.'s (2011) rationale, the same results were replicated, demonstrating that the lack of a bilingual advantage does not circumscribe to certain levels of education.

In Experiments 2A and 2B, a different approach was taken in order to delve into this same issue, this time testing a large group of bilinguals that differed in their L2 proficiency, ranging from a clearly low knowledge of the second language to perfectly fluent and balanced bilinguals. If bilingualism has any impact in cognitive functioning, then this impact should be modulated by the degree of knowledge of the second language, and thus the effects obtained in both tasks should show a significant correlation with seniors' L2 proficiency. Preceding research on this issue has failed at providing a consistent picture. Thus, to the best of our knowledge, the current study is the first one that aimed at checking for any possible modulation of bilingual seniors' inhibitory capacities by their degree of mastery of an L2 along a continuum, keeping all the other factors static. The 70 seniors tested in the second group of experiments came from the same city and all of them had acquired their second language before the age of 12, meaning that, despite individual differences in their use of the languages, the general degree of exposure to the languages in social contexts could be considered as highly homogeneous. The results from the ANCOVAs and correlations demonstrated that, regardless of their L2 proficiency participants showed comparable inhibitory skills (as measured by the Stroop effects), thus extending and qualifying earlier evidence obtained from younger bilinguals who did not show a significant relationship between L2 proficiency and the size of inhibitory control measures.

When the ANCOVAs were run including the IQ percentiles obtained from the K-BIT as a covariate, we replicated previous findings showing that the Stroop indices are reduced for higher IQ values (see Adelman, 2002; Arffa, 2007), but still we observed no main effect of Basque General Proficiency, nor an interaction of it with any of the indices. Similarly, when additional ANCOVAs were run including Education as a covariate, these same results were replicated. Participants' performance in the Stroop task improved as a function of Education (see also Houx, Jolles, & Vreeling, 1993; Moering, Schinka, Mortimer, & Graves, 2004; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006), but no effect of Basque General Proficiency was found, nor an interaction between Basque General Proficiency and any of the indices of interest.

Hence, the current results demonstrate that when the hypothesis of a potential modulation of inhibitory skills as a function of increased L2 proficiency is tested with lifelong elderly, no significant difference in their executive functions was observed as a function of their L2 knowledge.

The results found here add to the growing body of evidence that is gaining strength in the last several years and that suggest that the bilingual advantage in executive functioning (and explicitly in inhibitory and monitoring abilities) is actually non-existent. Different arguments have been used to reconcile the failures to obtain bilingual advantages. The absence of evidence favoring the bilingual advantage in young adults (Bialystok, Martin, & Viswanathan, 2005) has been argued to be a consequence of the ceiling effects that cognitive skills might feature in those ages; and therefore it might be captured easier in children. However, this has not been the case when contrasting a large sample of monolingual and bilingual children (Duñabeitia et al., 2014; Antón et al., 2014, Gathercole et al., 2014). In turn, this absence of evidence of an enhancement of general executive functioning in young children could also be argued to be due to the lack of enough exposure to bilingualism, meaning that these bilinguals have not undergone sufficient training in their lives. Consequently, it has been argued that the so-called bilingual advantage might emerge in later stages of life, given that the benefits of lifelong bilingualism could be better observed in samples of seniors whose cognitive skills are presumably declining. Nonetheless, here we demonstrated that in our elderly bilingual sample the bilingual advantage is absent. Furthermore, it is worth noting that our data is not the only one showing unambiguously that there are no differences in executive functioning between elderly monolinguals and bilinguals (Kirk et al., 2014; Kousaie & Phillips, 2012; de Bruin, Bak & Della Sala, 2015).

However, we acknowledge that the present results should be considered with cautiousness, and although our data show that bilingualism does not enhance executive functioning in the elderly, benefits derived from bilingualism in other domains should not be overlooked or disregarded, like the obvious benefits in terms of social and communicational skills. Nowadays, other potential paybacks of bilingualism at non-linguistic levels are also under debate, such as its neuro-protective value regarding the delay in the emergence of the symptoms of certain types of dementias (see Bialystok, Craik, & Freedman, 2007; Albán-González & Ortega-Campoverde, 2014). While some researchers do not support this potential consequence of bilingualism when the characteristics of the samples are carefully controlled for (see Chertkow et al., 2010; Lawton, Gasquoine, & Weimer, 2015), others report significant results even in carefully matched groups (see Alladi et al., 2013; Woumans et al., 2015). However, as correctly stated by Paap, Johnson and Sawi (2015b), the most compelling pieces of evidence at this regard may come from longitudinal studies following cohorts of individuals, and most of these studies yielded non-significant differences, or even monolingual-favoring trends (e.g., Crane et al., 2009; Lawton, Gasquoine, & Weimer, 2015; Sanders, Hall, Katz, & Lipton, 2012; Yeung, St. John, Menec, & Tyas, 2014; Zahodne, Schofield, Farrell, Stern, & Manly, 2014), while only one presented evidence in favor of a bilingual advantage at this level of analysis (Wilson et al., 2014).

The vast majority of research exploring the so-called bilingual advantage focuses on group comparisons, but conclusions from this type of experimental designs are always to be taken with caution, because despite all the effort put in matching samples, one might consider that dozens of factors can still play a role when the main comparison of interest is done based on a non-randomly distributed variable, i.e., bilingualism. In our present study investigating Basque-Spanish bilinguals and Spanish monolinguals, one could argue that Basque speakers represent a cultural minority with a different social and historical background from those of Spanish monolinguals. However, it is important to mention at this regard that our participants come from the same city and that they are not immigrants, thus sharing cultural and historical background to a great extent, and being as comparable as a between-subjects design allows for (Experiment 1). Furthermore, in an attempt to reduce the potential impact of uncontrolled between-groups factors, a second series of experiments were run considering the proficiency in the second language as a continuum in a sample of bilingual individuals (Experiment 2). Nonetheless, we acknowledge that other alternative approaches should be also favored. As suggested by Paap, Johnson and Sawi (2015b) and Duñabeitia and Carreiras (2015), it would be worth exploring how a wide range of cognitive skills (including executive functioning) changes before and after the acquisition of a language in the same group of individuals, following a longitudinal approach. According to recent evidence, it could be tentatively predicted that in case of the emergence of a difference (i.e., a bilingual advantage), this would be most clearly seen during the first years of immersion in an L2 context (see Heidlmayr, Moutier, Hemforth, Courtin, Tanzmeister, & Isel, 2014).

Conclusions

If we consider the evidence presented here together with the other published results showing no bilingual advantage in young children when critical confounding factors are controlled for, the argument that the ceiling effect of cognitive abilities is responsible for the lack of bilingual advantage in young adulthood is weakened. If that was the case, the advantage would be elusive in young adulthood but strong in children and the elderly. In addition, the present data (and in other studies cited above) showing no benefits of lifelong bilingualism weakens the hypothesis that there were no evident differences in children because lifelong exposure to bilingualism is needed to achieve that boosting effect in executive functioning. Thus, the evidence we collected in the present study adds to the growing body of evidence showing a comprehensive picture indicating that a bilingual advantage in tasks measuring executive control in any segment of the population is very likely to be produced by uncontrolled non-linguistic factors, rather than by the critical between-group difference of being bilingual or monolingual. As recently suggested, when those factors are controlled for and participant groups are carefully matched, no significant differences are captured between monolinguals and bilinguals in their performance in tasks tapping into inhibition and monitoring abilities (Paap, Johnson & Sawi, 2015a, 2015b). In fact, we failed to find any signs of a bilingual advantage in samples of carefully selected lifelong non-immigrant native bilingual seniors with different degrees of proficiency that live in a bilingual society, when their performance is compared with that of a group of carefully matched monolingual seniors. Hence, we conclude that lifelong bilingualism does not represent any specific benefit in executive functions in healthy elderly (see also Hilchey, Saint-Aubin, & Klein, 2015).

Acknowledgments:

This research has been partially funded by grants PSI2015-65689-P and SEV-2015-0490 from the Spanish Government, PI2015-1-27 from the Basque Government, AThEME-613465 from the European Union and an ERC-AdG-295362 grant from the European Research Council.

References:

Adelman, N., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., & Reiss, A. L. (2002). A developmental fMRI study of the Stroop color-word task. *NeuroImage, 16*, 61– 75. http://dx.doi.org/10.1006/nimg.2001.1046

Albán-González, G., & Ortega-Campoverde, T. (2014). Relationship between bilingualism and Alzheimer's. *Suma de Negocios*, *5*(11), 126-133. http://dx.doi.org/10.1016/S2215-910X(14)70027-8

Alladi, S., Bak, T. H., Duggirala, V., Surampudi,, B., Shailaja, M., Shukla, A. K., Chaudhuri, D. M., & Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology*, *8*, 1938-1944. http://dx.doi.org/10.1212/01.wnl.0000436620.33155.a4

Antón, E., Duñabeitia, J.A., Estévez, A., Hernández, J.A., Castillo, A., Fuentes, L.J., Davidson, D. J., & Carreiras, M., (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*, 398, 1-12. http://dx.doi.org/10.3389/fpsyg.2014.00398

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. Archives of Clinical *Neuropsychology*, *22*(8), 969-978. http://dx.doi.org/10.1016/j.acn.2007.08.001

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia, 17*, 467–472. http://dx.doi.org/10.1016/0028-3932(79)90053-8

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology*, *60*(1), 68-79. http://dx.doi.org/10.1037/cjep2006008

Bialystok, E., Abutalebi, J., Bak, T. H., Burke, D. M., & Kroll, J. F. (2016). Aging in two languages: Implications for public health. *Ageing Research Reviews*, *27*, 56-60. http://dx.doi.org/10.1016/j.arr.2016.03.003

Bialystok, E., Craik, F. I., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459-464. http://dx.doi.org/10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*(3), 89-129. http://dx.doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and aging, 19*(2), 290-303. http://dx.doi.org/10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F.I., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, memory, and cognition, 34*(4), 859-873. http://dx.doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for Mind and Brain. *Trends in Cognitive Sciences, 16*(4), 240–250. http://dx.doi.org/10.1016/j.tics.2012.03.001

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental science*, *7*(3), 325-339. http://dx.doi.org/10.1111/j.1467-7687.2004.00351.x

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103-119. http://dx.doi.org/10.1177/13670069050090010701

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science, 8*(6), 595-604. http://dx.doi.org/10.1111/j.1467-7687.2005.00451.x

Chertkow, H., Whitehead, V., Phillips, N., Wolfson, C., Atherton, J., & Bergman, H. (2010). Multilingualism (but not always bilingualism) delays the onset of Alzheimer disease: evidence from a bilingual community. *Alzheimer Disease & Associated Disorders, 24*(2), 118-125. http://dx.doi.org/10.1097/WAD.0b013e3181ca1221

Clark, J.L.D. (1981). Language. In T. S. Barrows et al. (Eds.), *College students' knowledge and beliefs: A survey of global understanding* (pp. 25-35). New Rochelle, NY: Change Magazine Press.

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*(2), 135-149. http://dx.doi.org/10.1016/j.cognition.2009.08.001

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*(1), 59-86. http://dx.doi.org/10.1016/j.cognition.2006.12.013

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*(5), 336-345. http://dx.doi.org/10.1038/nrn3709

Crane, P. K., Gibbons, L. E., Arani, K., Nguyen, V., Rhoads, K., McCurry, S. M., ... & White, L. (2009). Midlife use of written Japanese and protection from late life dementia. *Epidemiology, 20(*5), 766. http://dx.doi.org/10.1097/EDE.0b013e3181b09332

de Abreu, P. M. 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*(11), 1364-1371. http://dx.doi.org/10.1177/0956797612443836

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

Duñabeitia, J.A., Hernández, J.A., Antón, E., Macizo, P., Estévez, A., Fuentes, L.J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: myth or reality? *Experimental Psychology, 61* (3), 234-251. http://dx.doi.org/10.1027/1618-3169/a000243

Duñabeitia, J. A., & Carreiras, M. (2015). The bilingual advantage: Acta est fabula. *Cortex, 73,* 371-372. http://dx.doi.org/10.1016/j.cortex.2015.06.009

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics, 16*(1), 143-149. http://dx.doi.org/10.3758/BF03203267

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research, 12*(3), 189-198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

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

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., … & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in psychology*, *5*, 11. http://dx.doi.org/10.3389/fpsyg.2014.00011

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *The Journal of Neuroscience, 33*(2), 387-396. http://dx.doi.org/10.1523/JNEUROSCI.3837-12.2013

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition, 33*(7), 1220-1234. http://dx.doi.org/10.3758/BF03193224

Gollan, T. H., Salmon, D. P., Montoya, R. I., & Galasko, D. R. (2011). Degree of bilingualism predicts age of diagnosis of Alzheimer's disease in low-education but not in highly educated Hispanics. *Neuropsychologia, 49*(14), 3826-3830. http://dx.doi.org/10.1016/j.neuropsychologia.2011.09.041

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological science, 26*(4), 433-443. http://dx.doi.org/10.1177/0956797614567339

Heidlmayr, K., Moutier, S., Hemforth, B., Courtin, C., Tanzmeister, R., & Isel, F. (2014). Successive bilingualism and executive functions: The effect of second language use on inhibitory control in a behavioural Stroop Colour Word task. *Bilingualism: Language and Cognition, 17*(03), 630-645. http://dx.doi.org/10.1017/S1366728913000539

Heilenman, L. K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing, 7*(2), 174-201. http://dx.doi.org/10.1177/026553229000700204

Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastian-Galles, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition, 13*(03), 315-325. http://dx.doi.org/10.1017/S1366728909990010

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*(4), 625-658. http://dx.doi.org/10.3758/s13423-011-0116-7

Hilchey, M. D., Saint-Aubin, J., & Klein, R. M. (2015). Does bilingual exercise enhance cognitive fitness in nonlinguistic executive processing tasks. *Cambridge Handbook of Bilingual Processing.* Cambridge: CUP.

Houx, P. J., Jolles, J., & Vreeling, F. W. (1993). Stroop interference: aging effects assessed with the Stroop Color-Word Test. *Experimental aging research, 19*(3), 209-224. http://dx.doi.org/10.1080/03610739308253934

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production*?. Acta psychologica, 127*(2), 277-288. http://dx.doi.org/10.1016/j.actpsy.2007.06.003

Kaufman, A. S., & Kaufman, N. L. (1990). *Manual for the Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., Felber, S., & Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. *Neuroimage*, *25*(3), 888-898. http://dx.doi.org/10.1080/20445911.2014.929580

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology, 26*(6), 640-648. http://dx.doi.org/10.1080/20445911.2014.929580

Kousaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: Absence of a “bilingual advantage” in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology, 65*(2), 356-369. http://dx.doi.org/10.1080/17470218.2011.604788

Kroll, J. F., & Bialystok, E. (2013). Understanding the Consequences of Bilingualism for Language Processing and Cognition. J*ournal of Cognitive Psychology, 25*(5). http://dx.doi.org/10.1080/20445911.2013.799170

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition, 16*(03), 508-517. http://dx.doi.org/10.1017/S1366728912000508

Lawton, D. M., Gasquoine, P. G., & Weimer, A. A. (2015). Age of dementia diagnosis in community dwelling bilingual and monolingual Hispanic Americans. *Cortex*, *66*, 141-145. http://dx.doi.org/10.1016/j.cortex.2014.11.017

LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*, 673-687. http://dx.doi.org/10.2307/3586670

Lobo, A., Ezquerra, J., Gómez, F., Sala, J. y Seva, A. (1979) El mini-examen cognoscitivo. Un test sencillo y práctico para detectar alteraciones intelectuales en pacientes médicos. *Actas Luso Españolas de Neurología y Psiquiatría, 7*, 189-201.

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*(01), 81-93. http://dx.doi.org/10.1017/S1366728907003227

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child development*, *75*(5), 1373-1386. http://dx.doi.org/10.1111/j.1467-8624.2004.00746.x

Midgley, K. J., Holcomb, P. J., Van Heuven, W. J. B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research*, *1246*, 123–135. http://dx.doi.org/10.1016/j.brainres.2008.09.078

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*(1), 8-14. http://dx.doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100. http://dx.doi.org/10.1006/cogp.1999.0734

Moering, R.G., Schinka, J.A., Mortimer, J.A., & Graves, A.B. (2004). Normative data for elderly African Americans for the Stroop Color and Word Test. *Archives of Clinical Neuropsychology*, *19*, 61–71. http://dx.doi.org/10.1016/S0887-6177(02)00219-6

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental science, 10*(6), 719-726. http://dx.doi.org/10.1111/j.1467-7687.2007.00623.x

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science, 8*(1), 74-87. http://dx.doi.org/10.1111/j.1467-7687.2005.00394.x

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual 5 advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. http://dx.doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*(6), 615-639. http://dx.doi.org/10.1080/20445911.2014.944914

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

Paap, K.R., Johnson, H.A., & Sawi, O. (2015b). Should the search for bilingual advantages in executive functioning continue? *Cortex*, *74,* 305-314. http://dx.doi.org/10.1016/j.cortex.2015.09.010

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review, 16*(2), 225-237. http://dx.doi.org/10.3758/PBR.16.2.225

Sanders, A. E., Hall, C. B., Katz, M. J., & Lipton, R. B. (2012). Non-native language use and risk of incident dementia in the elderly. *Journal of Alzheimer's Disease, 29(*1), 99. doi: 10.3233/JAD-2011-111631

Santens, S., & Verguts, T. (2011). The size congruity effect: Is bigger always more? *Cognition, 118*(1), 94-110. http://dx.doi.org/10.1016/j.cognition.2010.10.014

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: the effect of an irrelevant cue on information processing. *Journal of applied psychology*, *51*(3), 300-304. http://dx.doi.org/10.1037/h0020586

Singh, N., & Mishra, R. K. (2013). Second language proficiency modulates conflict-monitoring in an oculomotor Stroop task: evidence from Hindi-English bilinguals. *Frontiers in psychology*, *4*, 322. http://dx.doi.org/10.3389/fpsyg.2013.00322

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643-662. http://dx.doi.org/10.1037/h0054651

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530-12535. http://dx.doi.org/10.1073/pnas.0609927104

Van der Elst, W., Van Boxtel, M.P.J., Van Breukelen, G.J.P., & Jolles, J. (2006). The Stroop Color-Word Test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, *13*, 62-79. http://dx.doi.org/10.1177/1073191105283427

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology an empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291-298. http://dx.doi.org/10.1177/1745691611406923

Woumans, E., Santens, P., Sieben, A., Versijpt, J., Stevens, M., & Duyck, W. (2015). Bilingualism delays clinical manifestation of Alzheimer's disease. Bilingualism: Language and *Cognition*, *18* (3), 568-574. http://dx.doi.org/10.1017/S136672891400087X

Yang, S., Yang, H., & Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism: Language and Cognition*, *14*(03), 412-422. http://dx.doi.org/10.1017/S1366728910000611

Yeung, C. M., St John, P. D., Menec, V., & Tyas, S. L. (2014). Is bilingualism associated with a lower risk of dementia in community-living older adults? Cross-sectional and prospective analyses. *Alzheimer disease and associated disorders*, *28*(4), 326. http://dx.doi.org/10.1097/WAD.0000000000000019

Zahodne, L. B., Schofield, P. W., Farrell, M. T., Stern, Y., & Manly, J. J. (2014). Bilingualism does not alter cognitive decline or dementia risk among Spanish-speaking immigrants. *Neuropsychology*, *28*(2), 238. http://dx.doi.org/10.1037/neu0000014

Appendix A: Tables

Table A.1. Characteristics of the samples of monolingual and bilingual seniors tested in Experiments 1A and 1B. Mean values for each group are displayed with standard deviations between brackets. T-tests report independent sample t-tests comparisons' results.

	Monolinguals		Bilinguals		p value
Chronological age (years)	68.75	(4.62)	69.38	(4.59)	0.64
Education (age)	15.58	(3.15)	16.17	(3.86)	0.57
MMSE (raw score)	29.13	(0.99)	29.17	(1.17)	0.89
General IQ (percentile)	59.67	(31.27)	65.33	(29.25)	0.52
Spanish general	8.54	(1.02)	8.67	(1.17)	0.69
Spanish speaking	8.67	(1.05)	8.54	(1.06)	0.68
Spanish comprehension	8.75	(0.9)	8.79	(1.06)	0.88
Basque general	--	--	8.04	(0.95)	--
Basque speaking	--	--	8.13	(1.08)	--
Basque comprehension	--	--	8.29	(1.37)	--

Table A.2. Mean latencies for correct responses and error rates in all conditions for the tasks tested in Experiments 1A and 1B for the monolingual and bilingual groups of seniors. Reaction times are showed in milliseconds with standard deviations between brackets. Errors are showed in percentages with standard deviations between brackets.

		Verbal Stroop task (mean RTs)				Numerical Stroop task (Mean RTs)			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	Congruent	772	(217)	787	(167)	615	(133.2)	608	(111)
	Incongruent	1017	(202)	1001	(176)	693	(153)	684	(133)
	Neutral Word	901	(185)	892	(131)	621	(125)	631	(111)
	Neutral Symbol	791	(144)	780	(105)	--	--	--	--
	Total	871	(167)	865	(130)	643	(131)	641	(114)
Effects	Stroop	246	(167)	213	(128)	77	(86)	76	(60)
	Congruency	129	(164)	104	(128)	5	(37)	24	(38)
	Incongruity	-117	(104)	-109	(125)	-72	(90)	-52	(73)

		Verbal Stroop task (mean error rates)				Numerical Stroop task (mean error rates)			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	Congruent	0.52	(1.41)	0.87	(1.73)	0	(0)	0.26	(1.28)
	Incongruent	2.26	(4.07)	2.95	(5.83)	2.08	(5.43)	2.34	(4.81)
	Neutral Word	0.69	(2.01)	0.69	(2.01)	0	(0)	0.52	(1.76)
	Neutral Symbol	0.35	(1.18)	0.17	(0.85)	--	--	--	--
	Total	0.95	(1.53)	1.17	(2.09)	0.69	(1.81)	1.04	(1.94)
Effects	Stroop	1.74	(4.06)	2.08	(5.63)	2.08	(5.43)	2.08	(4.76)
	Congruency	0.17	(2.6)	-0.17	(2.29)	0	(0)	0.26	(5.04)
	Incongruity	-1.56	(2.96)	-2.26	(4.91)	-2.08	(5.43)	-1.82	(1.28)

Table A.3. Characteristics of the sample of bilingual seniors tested in Experiments 2A and 2B. Mean values are displayed with standard deviation between brackets.

	Bilinguals	
Chronological age (years)	69.36	(4.4)
Education (age)	17.71	(4.71)
MMSE (raw score)	29	(1.3)
General IQ (percentile)	70	(29.65)
Spanish general	8.72	(1.08)
Spanish speaking	8.65	(1.07)
Spanish comprehension	8.99	(0.99)
Basque general	6.49	(2.4)
Basque speaking	6.7	(2.62)
Basque comprehension	7.23	(2.13)

Table A.4. Mean latencies and error rates for the bilinguals seniors tested in Experiments 2A and 2B. Reaction times are presented in milliseconds and error rates in percentages. Standard deviations are displayed between parentheses.

		Verbal Stroop task				Numerical Stroop task			
		Reaction Times		Error Rates		Reaction Times		Error Rates	
Conditions	Congruent	789	(145)	0.42	(1.26)	598	(103)	0.45	(1.62)
	Incongruent	1011	(174)	2.14	(4.42)	671	(117)	1.79	(4.00)
	Neutral Word	891	(140)	0.36	(1.37)	621	(100)	0.27	(1.28)
	Neutral Symbol	765	(104)	0.18	(0.85)	--	--	--	--
	Total	864	(128)	0.77	(1.46)	630	(103)	0.83	(1.60)
Effects	Stroop	223	(107)	1.726	(4.29)	73	(56)	1.340	(4.37)
	Incongruity	-120	(98)	-1.79	(4.11)	-50	(61)	-1.52	(3.90)

Congruency	103	(108)	-0.06	(1.66)	24	(34)	-0.18	(1.83)

Appendix B: Figures

Figure B.1: Correlations in the verbal Stroop task. Correlations between *Stroop* (Incongruent-Congruent), *congruency* (Neutral Words-Congruent) and *incongruity* (Neutral Words-Incongruent) indices in latencies (in the vertical axis, indicated as the difference between conditions in milliseconds) and General Basque Proficiency (in the horizontal axis, from 1 to 10) for the Experiment 2A (verbal Stroop task). R^2 value is indicated in each graph.

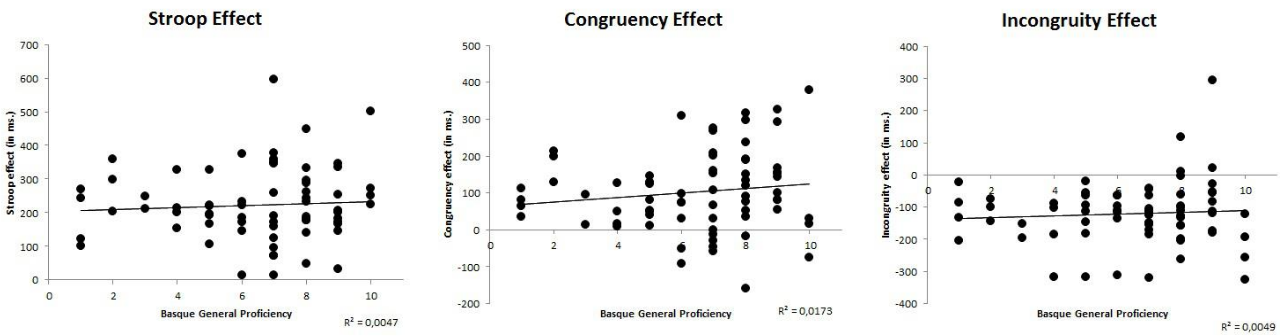

Figure B.2: Correlations in the numerical Stroop task. Correlations between *Stroop* (Incongruent-Congruent), *congruency* (Neutral-Congruent) and *incongruity* (Neutral-Incongruent) indices in latencies (in the vertical axis, indicated as the difference between conditions in milliseconds) and General Basque Proficiency (in the horizontal axis, from 1 to 10) for the Experiment 2B (numerical Stroop task). R^2 value is indicated in each graph.