The Search for a Bilingual Advantage in Executive Functions: a Developmental Perspective

Eneko Antón Ustaritz

2017

BASQUE CENTER
ON COGNITION, BRAIN
AND LANGUAGE

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

The Search for a Bilingual Advantage in Executive Functions: a Developmental Perspective

Doctoral dissertation by

Eneko Antón Ustaritz

In candidacy for the Degree of Doctor in Linguistics

Supervised by:

Dr. Jon Andoni Duñabeitia and Prof. Manuel Carreiras

Donostia-San Sebastián, 2017

The research presented in this thesis was supported by grants PRE_2013_1_1155, PRE_2014_2_259, PRE_2015_2_0167 and PRE_2016_2_0065, from the Department of Education, Linguistic Policy and Culture of the Basque Government.

BCBL Basque Center on Cognition, Brain and Language

Paseo Mikeletegi, 69,

Donsotia-San Sebastián, Spain

March 2017

Aitonari

Eta Bojanari

Acknowledgements

Tesi hau pertsona bakar batek idatzia izan baldin bada ere, askoren indar eta maitasunak kementsu bultzatu eta lagundu izanaren ondorio zuzena da. Ziur aski ez dira hemen egon behar luketen izen guztiak agertuko. Barkatzen jakingo duzuelakoan, zuei ere eskerrak.

De entre todos los colegas, compañeros y trabajadores del BCBL, me gustaría dar un agradecimiento especial a Jon Andoni y a Manolo por haber supervisado esta tesis. Pero sobre todo, por ser mentores y guías durante estos años en los que es fácil perderse. Gracias sobre todo a Jon Andoni por enseñarme el valor del pensamiento crítico y por instruirme para ser parte de este mundo, dejándome tomar parte activamente en su equipo y recompensando el trabajo duro.

Lanetik kanpo, noski, familiarentzan esker haundiena, aita, ama, aitona (goian bego), amona, izeba Edurne eta izeba Marta. Zer egiten dudan oso ondo jakin gabe ere, beti animoso, laguntzaile eta harro, aurrera egitera erakutsi didatelako.

My greatest gratitude also goes to Bojana, *oko moje,* for being more than a partner and a colleague, for deciding to walk with me in this complicated journey, for being caring, supportive and understanding both in the academic and private life. *Hvala puno na ljubavi i podršci, volim te.* Now I'll try to do the same for you, I'm sure you'll to rock it.

Gracias también a todos los integrantes de La Cuadrilla, porque las cervezas de después del trabajo o los fines de semana ayudan más de lo que parece para afrontar un doctorado con fuerzas. En especial a Lasa (¡o Dr. Lasaosa!) porque no hay plan al que no se apunte ni debate que no quiera llevar al más absurdo extremo. El mejor entrenamiento para defender mis argumentos en conferencias y revistas.

Tampoco me olvido de la cuadrilla del trabajo, sobre todo de la comunidad predoc. Empezando por los que estuvieron desde el principio (Saioa, Aina, Alejandro, Mikel, Karla, Cris…), los que fueron viniendo después (Ainhoa, Paula, Saúl, Irene, Gari, Alexia, Patricia, Lela, Borja, Noemí, Asier, Joyse…) y los que nos visitaron sólo durante un tiempo (Francesco, mate, *grazie mille*). Un agradecimiento muy fuerte y especial para Myriam, por ser la alegría y la guía espiritual de todos, and also to Peter, for this bromance grounded on legendary puns. You are the milk. My sincere gratitude to him and to Brendan for their marvelous and uninterested help during the process of reviewing this thesis. Last, but not least, a very special *hvala puno* to Jovana. *Ti si do jaja.*

Por supuesto, un agradecimiento muy fuerte también a todo el personal no investigador del BCBL, administración (en especial a Eider, por su buena voluntad y ayuda en estos últimos compases), personal de laboratorio y IT, ya que sin ellos este trabajo no hubiera sido posible. Gracias también a Pilar, por habernos ayudado a todos a conectar un poco mejor con la realidad.

Pertsona guzti hauen laguntzarik gabe, tesi hau ez litzake izango. Eta izango balitz, guztiz desberdina izango litzake. Beraz, hemen goian izendatu zaituzten guzti-guztioi, eta izendatu ez zaituztenoi ere, eskerrik asko, muchas gracias, thank you so much, hvala puno, bihotz-bihotzez.

Abstract

Bilinguals need intensive language-control mechanisms to produce effective communication and avoid intrusions from the non-target language, because both languages are always active in a bilingual mind (Thierry & Wu, 2007), competing with each other. It is mostly assumed that bilinguals apply inhibition to the non-target language (e.g., the IC model, Green, 1998). The bilingual advantage theory claims that this constant need of inhibition trains bilinguals' general inhibitory abilities, making them better than their monolingual counterparts in any situation where inhibition is needed (Bialystok et al., 2005). However, it has been recently argued that the repeatedly shown bilingual advantage effect in tasks tapping into domain general inhibition might stem from uncontrolled factors associated to bilingualism, rather than from bilingualism itself, as well as from small sample sizes (Paap & Greenberg, 2013). Crucially, previous evidence tended to neglect the importance of factors that correlate with better executive function abilities, such as immigrant status or socio economic status (Mezzacappa, 2004; Milne, Poulton, Caspi, & Moffitt, 2001). In this thesis I aimed at testing the reliability of the bilingual advantage by testing large samples of bilingual and monolingual participants of different ages, matched in the relevant extra-linguistic factors. Bilingual and monolingual children, young adults, and seniors underwent several classic tasks that tap into domain general executive functions, such as the verbal and numerical Stroop task, the Flanker task, and the Simon task. If the bilingual advantage exists independently of the previously uncontrolled confounding factors, bilinguals should show a reduced conflict effect in (at least) some of the abovementioned tasks, reflecting a better ability to inhibit irrelevant information. No indication of any bilingual advantage was found whatsoever. The bilingual and monolingual groups behaved comparably in every task, obtaining highly similar indices. The results are discussed and interpreted in the light of different perspectives, mainly questioning the origins of the bilingual advantage theory, which claims that the cognitive mechanisms responsible for domain general inhibition are also responsible for language control. However, as no correlation was found between the indices that the participants obtained in the different general inhibition tasks tested here, it is proposed that bilingualism might arguably enhance inhibitory abilities required in language control only, and not the ones required in domain-general situations.

Table of Contents

Chapter 1: General introduction

In modern society, bilingualism is more the rule than the exception. Countries with more than one official language, teaching programs in two languages and increasing mobility between countries have introduced bilingualism in almost all strata of the society. It is not surprising to observe that current estimates situate the percentage of bilinguals at more than half of the planet (Grosjean, 2010). Furthermore, the proportion of bilinguals is increasing (Bhatia & Ritchie, 2008), as well as the number of people who speak a language other than the official one at home (Shin & Kominski, 2010). Many social factors contribute to this growing amount of bilinguals but, importantly, government-promoted changes with new bilingual educational policies have stimulated this situation, with some estimates indicating that around two thirds of children in the world are raised in a bilingual environment (Crystal, 1997).

Bilinguals, by definition, are people who *"speak two languages fluently"* (Oxford dictionary). They are very efficient in their communicative skills, and they generally achieve successful communication with little or no cross-language contamination (unless code-switching is intended, see Deuchar, 2005). The fact that bilinguals can freely choose which language to use according to the communicative situation is a well-accepted fact, but bearing in mind how demanding the processes involved are, one could deem such success surprising. No matter how efficiently bilinguals can use one language or the other and achieve successful communication with little apparent effort, a large amount of evidence coming from research undoubtedly shows that both languages that a bilingual speaks are always active (Kroll, Dussias, Bogulski, & Kroff, 2012; van Hell & Dijkstra, 2002). How does a bilingual manage to efficiently choose between the two languages? How is this process controlled in such an efficient way? Some researchers have argued for the existence of special control mechanisms in bilingual minds that allow for efficient language management (Kroll et al., 2012; Green, 1998), which could be different from those control mechanisms that are present in monolinguals. As a consequence of this, some authors have argued that the use of these special control mechanisms by bilinguals could create a training transfer, that is to say, the skills that bilinguals show in using one language and avoiding (or inhibiting) the non-target language might be extended to any other general cognitive ability. Hence, this training transfer could apparently enhance domain general control abilities (Bialystok, 1999; Bialystok & Martin, 2004), which would reciprocally account for the efficiency in bilinguals' communicative skills.

I would like to re-examine these assumptions and see whether these special mechanisms truly exist in bilinguals. Under the assumption that the two languages in bilinguals interact or affect each other, it could be expected that the bilinguals would indeed need some control mechanisms. Yet, there is an ongoing debate on how the languages are organized in a bilingual mind as compared to a monolingual's, and whether or not the two languages affect each other in any way. I will try to shed some light on those issues, and, importantly, on whether there is any linguistic or cognitive consequence, either positive or negative, of being a bilingual speaker.

In this first chapter I present an overview of the classic and current perspectives on language organization in monolinguals and in bilinguals, to see how those classic models and paradigms can apply to the specific case of a person who has two (or more) language systems. Later on in this chapter, I will focus on the critical issue of how bilinguals are able to deal with two languages efficiently, explaining the history of evolution of the main views on this topic. After explaining how bilinguals operate, I will explore the linguistic consequences of having two languages instead of one, to later move to the potential cognitive differences produced by bilingualism that have been reported in the literature. These linguistic and cognitive differences have been shown to be positive sometimes, and also negative in other cases, and there is a solid consensus for some of them but a lively debate for others. Among all these differences, I will dedicate more space and time to the main construct explored in this thesis: the so called *"bilingual advantage in executive functions"*. For the sake of simplicity and clarity, and although other advantages have been reported, every time I refer to the *"bilingual advantage"* hypothesis in the present thesis I will be referring to the advantage that some authors report that bilinguals show in executive functioning as compared to monolinguals. This term has been used in the past years to refer to a supposedly better performance that bilinguals show in tasks requiring executive control, and especially inhibition, when they are compared to monolinguals. I will review the perspective and evidence supporting it (e.g., Bialystok, 2006) and the criticisms that it has received (e.g., Paap & Greenberg, 2013). In order to explore whether the criticisms to the bilingual advantage are coherent and have a solid grounding, I will present data from the experiments that were conducted to explore the bilingual advantage in seniors, children and young adults (second, third, and fourth chapters, respectively). These experiments used the same tasks that previous research that has reported to show significant effects of bilingualism used but, critically, the criticisms against and limitations of this earlier work that may have led to specious significant results in the past were taken into account. The results I find and their implications for the bilingual advantage debate are explored in the final chapter, the general conclusions.

I. *Language organization in monolinguals: The mental lexicon and lexical access*

To understand how language is organized in our minds, it is useful to first have a brief look to a broader picture and to try to understand how the mind itself is organized, with its particular functions and specific mechanisms. A ground-shaking publication in this respect was Fodor's *"The Modularity of Mind"* (Fodor, 1983). Although the purpose of the present thesis is not to discuss the concept of *"modularity"* (Fodor, 1983; 2001) in depth, it is a construct worth clarifying given its implications for the upcoming paragraphs. According to Fodor's modularity perspective, the human mind is organized in three levels of cognitive processing (see Fig. 1): the transducers, which convert physical information into neural signs that can be processed and worked with; the input modules, responsible for basic cognitive functions that interpret the information coming from the transducers (which would be aligned with behaviorist perspectives of cognition); and the central system, which is responsible for higher cognitive abilities (such as reasoning) and is not modular. The cognitive systems that are organized in a modular fashion in the input system are domain specific, obligatory, innate, fast and specific (which

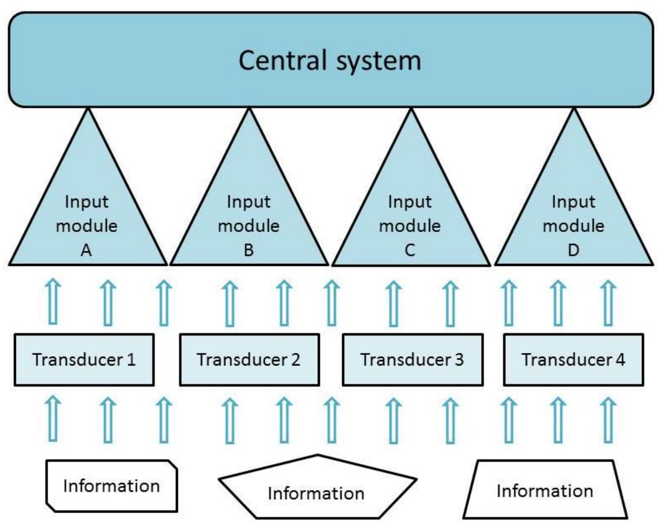

Figure 1. Schematic representation of the modularity of mind. Adapted from Fodor (1983)

sometimes is equated to neural specificity). The information is encapsulated (meaning that these modules are unaffected by other cognitive domains and other modules) and they have a limited access to the central system. What all these characteristics mean, in combination, is that these modules respond innately, quickly and automatically to a specific set of external input and only to that; and that they are not modulated by and cannot access to other modules. These sets of external input that the modules respond to are called domains, and they are much more fine-grained than sensory modalities, comprehending stimuli such as faces, visual objects, spoken language, etc.

This theory received much criticism and opposition from both theoretical and experimental point of view (Churchland, 1988; Arbib, 1987; Marslen-Wilson & Tyler, 1987; McCauley & Henrich, 2006; Hulme & Snowling, 1992; Bishop, 1997; Wojciulik, Kanwisher, & Driver, 1998; Prinz, 2006), leading to updated versions of the modularity approach. For example, Carruthers (2006) defended the thesis of a *massive modularity*, arguing (contra Fodor's module conception) that the central system is also modular, but in a weaker way than the input systems (see Wilson, 2008; for a response on this issue). The general principles of modularity have also been applied to language. In this perspective, language – and particularly each of its different forms (spoken language recognition, written language recognition, etc.) – would have its own modules that would respond only to those specific inputs. It has been proposed (Block, 1995) that language (i.e., each language module), like any other module, should have an internal structure of a set of smaller modules, decomposed into primitive processors such as the phonetic awareness module, the syntax module, etc. This idea was to some extent suggested by Fodor himself when he proposed the "within module interlevels". It basically means that any domain input module is formed by interlevels that interact with each other. For example, in the visual-object recognition module, different levels would play different roles and some would analyze the visual inputs, others would understand the 3D sketch behind it and others would pair that 3D sketch to a basic category in the form-concept dictionary. In language, the same principle would be applied. For example, in the spoken-language module, the lexical and the phonetic processor levels could interact and help each other (thus explaining different language phenomena such as "phonetic restoration", Warren, 1970). Similarly, in the written-language module, the alphabetic letter representation level would account for the recognition of the same letter in different shapes (A-a), and the level of shape-specific letter representation would identify letters by finding a match in the shape-specific level.

While this concept of modularity and the internal relation between modules is useful to understand the upcoming models, Fodor's theory was not intended to explain only language, but it was applied to the whole organization of cognitive skills. Besides Fodor's model, as far as language is concerned, concerted efforts have been made for decades to try to explain how this cognitive ability in particular is organized and processed. Concretely, these studies have classically focused on how words are stored and how we access them. In the upcoming sections I will try to explain the different perspectives that have been adopted throughout the years and the agreements that have been reached in those regards.

1. *The mental lexicon*

One of the biggest classical questions that psycholinguists have tried to answer is "How are words organized in our minds?" Words are assumed to be organized in a coherent way, which allows us to access them efficiently to either recognize or produce them. The *"mental lexicon"* is the name with which the systematic organization of vocabulary in the mind is known, and it contains the individual lexical entries that a particular person knows (Field, 2004). According to Levelt (1989), these lexical items contain two kinds of information that allow people to identify and understand words: the form (morphological and phonological information) and the meaning (semantic information). The access to the lexical entries stored in the lexicon has been described many times as the cognitive equivalent to looking for a word in a dictionary. However, the organization and access of the words in the mental lexicon is much more complex and multi-faceted given the different features based on which lexical items are organized and accessed, and not only by their alphabetically ordered spelling as occurs in dictionaries (Fellbaum, 1998).

Unfortunately, the way in which words are organized in the mental lexicon has been object of debate. Different perspectives and models have been proposed and have evolved over the years. Originally, it was believed that the organization was *hierarchical* (see, for example, Collins & Quillian's *hierarchical network model,* 1969), with words organized in a sort of pyramidal structure where the most general items are on the top (e.g., "animal") and the more specific entries are organized in individual nodes in steps to the bottom (from "animal" down to "bird", then to "canary"…). Soon the shortcomings of these models were evident and *semantic feature models* were developed (e.g., Smith, Shoben, & Rips, 1974). These models emphasize the importance of words' semantic features and argue that lexical entries are organized based on them. They differentiate between words' defining features (essential for the meaning of the word, the most salient features) and characteristics (non-essential features). Thus, concepts that share many defining features would be stored close to each other with a strong relation between them. For example, different kinds of birds would share "has wings" and "has a beak" and they would be stored together, but among all the birds some would be more closely stored than others: robin and eagle share "has wings", "has a beak" and "flies", whereas robin and ostrich only share the first two. Robin and eagle would be closer than robin and ostrich in this model, while in the hierarchical model the three of them would be equally close to each other and to the superordinate word "bird". Later, Collin and Loftus (1975) tried to adapt the hierarchical network model (Collins & Quillian, 1969) to the criticisms received, and they proposed a model that somehow was similar to semantic feature models, the *spreading activation model* (Collins & Loftus, 1975). However, their model did not rely on feature comparison to group lexical entries together, but treated objects, features and verbs as different concepts in separate nodes or entries. The words (objects, features and verbs) can be interconnected with different degrees of strength (see Fig. 2). Importantly, this model states that when a node is activated, that activation spreads to the adjacent nodes, activating them as well (the spread activation principle). That pulse of activation that spreads to related words weakens over the links until it finally dissipates, so the closer two concepts are, the more activation one will receive when the other is activated. This co-activation

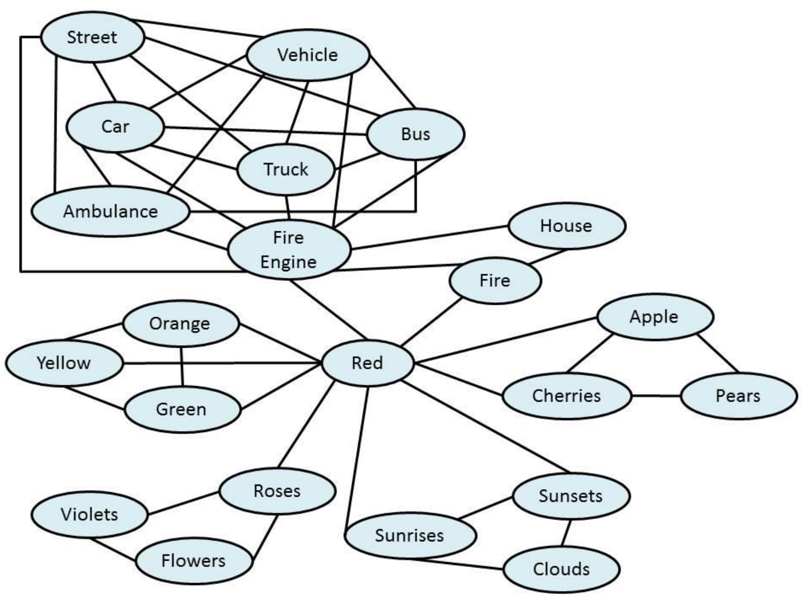

Figure 2. Representation of lexical items arranged according to the spreading activation model. Adapted from Collins & Loftus (1975).

explains why priming occurs: it is easier to recognize a target word when a semantically related one has been presented previously and thus, all the words stored closely (including the target word) have been activated (Neely, 1977).

Different models not only offer different structure for the organization of the mental lexicon, but they also differ in explaining how speakers can access lexical items that are stored in the lexicon upon encountering them visually or auditorily. Some authors (e.g., Forster's *autonomous search model*, 1976) argue that the processing of language is *serial*: we encounter a word, we compare it to our different lexical entries and, if we find a correspondence, we retrieve the necessary information about it. Then we move to the next lexical entry. *Parallel access models*, in contrast, argue for parallel activation of multiple entries when the perceptual input of a word happens. Potential candidates are activated and the one that shares the most features with the target item is the one selected (Gleason & Bernstein, 1998). Currently these parallel models are mostly accepted over the previous serial processing models. Some of the most studied models are Marslen-Wilson's (1987) *cohort model* and Seidenberg and McClelland's (1989) *connectionist models*. The cohort model assumes that, when an individual hears a word, all the phonological neighbors (i.e., the *cohort*) are activated. This is a similar concept to the one introduced above for the spreading activation model (Collins & Loftus; 1975), with the exception that the co-activation is phonologically, rather than semantically driven. Thus, when the word's first phonemes are perceived, every word that shares those few sounds from the beginning (the *cohort*) gets activated. By narrowing the cohort, with more phonetic input, all the activated words are progressively eliminated until a unique lexical entry remains. This explains why a phonologically similar prime facilitates the recognition of a target word (Ferrand & Grainger, 1992).

The models of lexical organization and access are still developing with more updates from classic models to new proposals based on new discoveries. However, for the sake of simplicity and considering the purpose of this thesis, I adopt a condensed and summarized vision of the various perspectives on how the mental lexicon is organized, and consider that it is made in such a way that words are connected with each other based on many features, from semantics to phonology. Thus, access to or activation of one of the words/nodes would involve spreading activation to adjacent nodes (following the spreading activation principle), activating them and resulting in a set of potential candidates competing for selection. In a successful setting of communication, the correct and intended candidate would be chosen.

2. Accessing the mental lexicon to produce words

Similarly to lexical organization and access, there is no conclusive model on language production that brings different theories to a consensus. The question of how concepts are selected to be produced in the actual utterance still needs to be clarified, and the question of how a concept is translated into an actual utterance once selected is a topic of debate. Although language production has been one of the central topics in psycholinguistic research since its foundation, after decades of intense research, there is no full agreement on one single model that explains the process of producing speech. In a broad sense, and based on the characteristics that different models propose, models of language production can be grouped into *serial processing* and *parallel processing* models.

In the first group, the *serial* or *discrete* models (e.g. Levelt, 1989; Levelt et al., 1991; Levelt, Roelofs, & Meyer, 1999), one of the most accepted and studied models is that originally proposed by Bock and Levelt (1994). In their model, they propose that firstly the speaker chooses the message he wants to transmit, the main idea to be conveyed, in form of concepts. Once the idea is clear, in the functional level the concept representation that the speaker had in mind is turned to a lexical representation by the selection of words (in the lexical selection stage) and by assigning a syntactic function to them (in the function assignment stage). In the third level, called the positional level, the inflection and the order of the

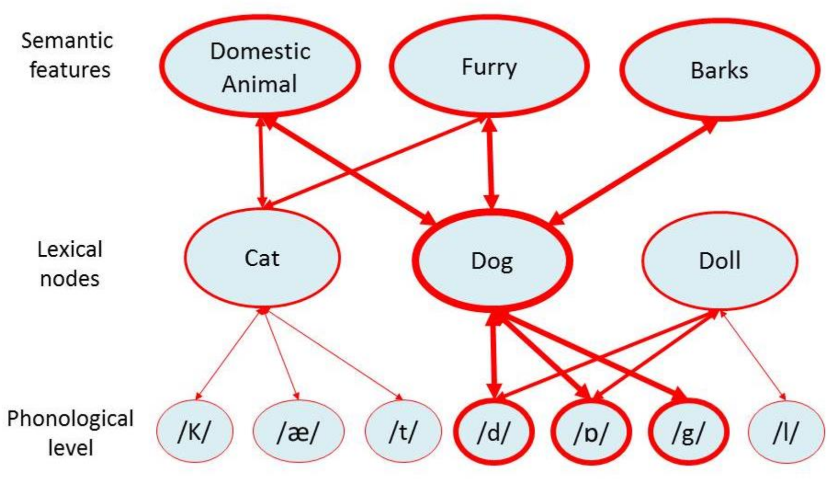

Figure 3. Representation of the parallel processing activation. The diagram shows the spread of activation produced by the word “dog”. The thickness of the arrows and nodes represents the strength of the activation. Adapted from Dell (1986).

morphological slots are determined. This allows the fourth phonological encoding level to assemble phonemes and their intonation into lexemes, based on words' phonological and morphological properties. Finally, the product of the whole process is sent to the articulatory system, the output in which the actual speech is produced.

When it comes to the *parallel processing* models (see, for example, Starreveld & La Heij, 1995, 1996; Caramazza, 1997; Rapp & Goldrick, 2000; Navarrete & Costa, 2005), one of the best representatives is that proposed by Dell (1986; see also Dell, Chang, & Griffin, 1999). This model goes against the hierarchy proposed by the serial models, and claims that language is produced by means of several different and interconnected nodes, each one representing a different level of language (phonemes, morphemes, syllables, concepts, etc.) that can interact with each other in any direction. This model encompasses three main levels, namely the semantic features level, lexical nodes level and phonological structure level, with connection paths among all of them (see Fig. 3). This again introduces a concept similar to the spread activation principle. It means that when, for example, the concept of "dog" is to be produced, the nodes carrying the semantic, syntactic, phonetic and morphological features of that word are activated, and that activation is spread to the nodes to which they are connected. In the example, the concept "dog" would activate the semantic features "domestic animal" and "furry", which probably would also activate the concept of "cat". Furthermore, after the lexical item for "dog" is activated and then the activation is spread to the phonetic level, each of the phonemes of the intended word "dog" would also activate adjacent words that share those phonemes, such as "doll". In contrast to the serial models explained above, where the activation flow is restricted from only the selected lexical representation to its phonological content, in these parallel processing models the activation spreads both forward and backwards in the whole system (e.g. Dell, 1986; Dell & O'Seaghdha, 1991; Rapp & Goldrick, 2000). Importantly, the spread activation would flow not only from the target lexical item ("dog") to its phonological representation and back, but also from the closely activated lexical items ("cat") to their corresponding phonological nodes (Costa, Caramazza, & Sebastián-Gallés, 2000). This means that the lexical nodes within each level compete for selection based on activation, and in the end the most activated one is the one to be produced. This spread of activation to the adjacent competing nodes explains that sometimes the wrong word is picked and produced, given the strong activation of the nodes that are semantically or phonetically very similar to the target. That situation needs to be avoided if successful communication is to be reached, and that is why, as later argued by Dell and O'Seaghdha (1994), language nodes not only compete for activation, but also send inhibitory signals to one another and apply self-inhibition if needed. Inhibition was first considered for perception models (see, for example, the *interactive activation model*, McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) with great success, and later applied to production models also (see Berg & Schade, 1992; Harley, 1990; or Schade & Berg, 1992; among many others). Along those lines, Dell and O'Seaghdha (1994) argue that, in the activation-based production models such the ones just mentioned, excitatory and inhibitory inputs are sent between units, modulating the activation level. For example, competing nodes within the same category send lateral inhibition to each other, especially if they represent mutually exclusive representations. Verbs such as "give" and "donate" would inhibit each other since only one of them can be used, but also "give" would activate some syntactic features (it needs two syntactic objects in the sentences) and inhibit others (a prepositional dative configuration that would be used with donate). Furthermore, semantically similar candidates would be also inhibited. Thus, if we want to produce a concept that is within the category of the things that are used to drink liquids, the competing "mug" should be strongly inhibited for the target "cup" to be produced. Neighbors would be thus roughly inhibited in proportion to their relatedness, with a stronger inhibition applied to stronger candidates that are more likely to disrupt the ongoing speech (for a more detailed explanation of inhibitory processes, see Dell & O'Seaghdha, 1994).

Despite their differences concerning the way in which the linguistic units are selected, arranged and produced, it can be observed that most of the models have the same basic ideas of how language works. For the sake of simplicity, I assume that speech production works following the most accepted and general steps: first the concepts are conceived, these are turned into words and, if needed, arranged in a sentence, and finally they are produced. As was explained for the organization of the mental lexicon, the activation of each level (semantics, orthography, phonology…) affects the others (the abovementioned spreading activation principle, Collins & Loftus, 1975) in top-down and lateral connections, and these connections are regulated in different ways via excitatory and, crucially, inhibitory inputs sent to the adjacent nodes. All the nodes are interconnected and thus the activation is spread and inhibition needs to be managed to block the competing candidates through all the levels.

It is worth noting that, no matter what interpretation different authors have given to where and how they locate the various mechanisms for lexical access and speech production, all these models and proposals have only focused on the essential linguistic variables such as semantics, syntax and phonetic activation. The abovementioned classic models did not explore in depth the extent to which having two language systems would significantly affect the mental lexicon and its functioning. The questions regarding this issue concern the way in which the second language relates to the first one, how the spread activation principle is applied to the second language (or whether it is or not), as well as the existence of potential inhibition between the two languages. A further issue is whether the words of both languages constitute plausible candidates for selection during utilization of only one language, or the languages act independently. I will try to address these issues in the next paragraphs.

II. Language organization in bilinguals: The mental lexicon and lexical access

Based on the models described in the previous section, it seems that we are far from reaching agreement on defining how language is organized in the mental lexicon and how it is accessed. Most certainly, there have been considerably fewer attempts to answer these same questions in the case of bilinguals. Different models and paradigms have been developed to try to understand the role that the two languages play and how they affect to each other.

One point on which most lexical access models concur with respect to the role of bilingualism in the system is that the semantic system is shared between the two languages of a bilingual speaker (De Bot, 1992; Costa, Miozzo, & Caramazza, 1999; Kroll & Stewart, 1994; Poulisse & Bongaerts, 1994; Dijkstra & van Heuven, 1998; 2002; French & Jaquet, 2004; however, some researchers have proposed that semantics are language specific, see Van Hell & De Groot, 1998). Beyond the semantic domain, the next question is whether or not the semantic activation spreads to the lexical level of the target language only, or to both languages that a bilingual speaks. If the activation that originated at the concept level spreads with no language restriction, then a bilingual speaker would have many more competitors to the target word than a monolingual speaker. For example, if a monolingual is presented with a picture of a dog, and she has to say "dog", several steps have to occur for that name to be uttered. As described in the previous section, the concept of the dog would be activated first, in the semantic representation, and the activation would spread to the adjacent semantic nodes based on either semantic proximity ("cat", "fish", etc, according to models such as Levelt, 1989) or on shared semantic features ("has four legs" and "is a domestic animal" according to models such as Dell, 1986). Regardless of the processes through which the activation spreads, the semantic neighbors of the word would be activated when the target word needs to be produced. The spread of the activation would continue, and the adjacent semantic nodes that were activated would now spread activation to their corresponding lexical labels in the lexical level. At this point, several lexical items are activated and competing for selection (which can be taken as the source of semantic mistakes in speech production, see Caramazza & Hillis, 1990), and the appropriate one needs to be chosen. Generally, the most activated item is the one chosen to be produced. In this context, the more activated the competitors are, the more difficult it would be to pick the correct word. If the most activated item is not the target word, the speaker would produce an error (e.g., "the cat is barking"). Bilinguals, however, would activate not only "dog" and its closely related ones, but the target word in the language not in use ("perro", the Spanish for "dog") and its adjacent entries would be activated too, and they would all compete for selection. If that is so, bilingual lexical access could be much more effortful than monolinguals lexical access and it might require some control mechanisms and more fine grained abilities to avoid the conflicting information from the non-target competing lexical items.

There is debate on the existence of parallel activation of languages (Costa, La Heij, & Navarrete, 2006). As mentioned in the beginning of this section, the semantic system is shared among the two languages of a bilingual speaker (Costa, Miozzo, & Caramazza, 1999; Kroll & Stewart, 1994; Dijkstra & van Heuven, 1998; 2002). Hence, the common semantic level would be connected to the lexical nodes of each of the languages. The question that arises is whether the spreading of the activation is language specific or not, i.e. whether or not the activation on the semantic nodes spreads to the lexical items of both languages or just to the target one. Some perspectives have argued for some kind of *"switch device"* that would allow the speaker to turn on or off the lexicons of the different languages so that only the target language is active (McNamara & Kushnir, 1972; Macnamara, Krauthammer, & Bolgar, 1968). Those classic explanations are no longer considered, and the two main approaches currently in the bilingual lexical access field are the *language selective* perspective, which maintains that language is selected before lexical access and thus only the target language's candidates will be

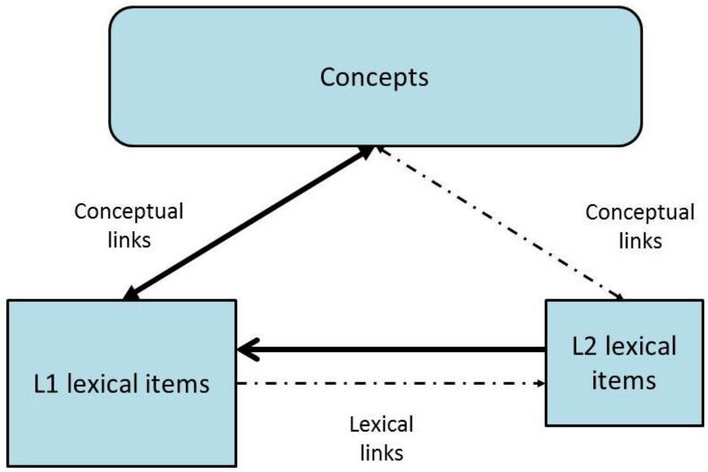

Figure 4. Representation of the RHM. The thickness of the arrows represents the strength of the connection between the nodes. Adapted from Kroll & Stewart (1994).

activated, and the *language non selective* access hypotheses, which argues in favour of an activation of both the target and the non-target language, which implies a subsequent competition between languages (Dijkstra, 2005). This difference seems crucial for the impact that bilingualism might have on cognition, because in the second case, some control mechanisms should exist to avoid cross-linguistic interference. These control mechanisms would not be necessary in a situation in which lexical activation is restricted to the intended language. If the access is *language selective*, the language would be selected before the recognition of a word and therefore only the target language system would be activated (Grainger & Beauvillain, 1987), with only the lexical nodes of the language in use being activated (Costa, Miozzo, & Caramazza, 1999; Costa & Caramazza, 1999). In this case, the non-target language's lexical nodes would be ignored and no control mechanisms would be needed.

In a case in which the access is *not language selective*, both language systems would be interconnected and activated (Kroll & Tokowicz, 2005). Reflecting the connection between the two language systems, Kroll and Stewart (1994) postulated in their r*evised hierarchical model* (RHM) that when the L2 is acquired, its lexical items are assimilated through L1 translation. In the early stages of becoming a bilingual (see Fig. 4), the two translation equivalents would be directly interconnected, although the new L2 items would also have a direct connection to the conceptual level, which is assumed to be weaker than the connection to their L1 translation and the connection between L1 items and concepts. Thus, when an L2 word is encountered, it would be understood (i.e., access to concepts) firstly through its L1 translation, reflecting the importance of the between-language links. The strength of the direct link from L2 lexical entries to the concept node depends on the proficiency that the individual has in the languages. The level of proficiency determines whether access to the L2 word meaning is straightforward or mediated by L1 translation. For example, a highly proficient bilingual would not need to access L2 word meaning through L1 translation, since his L2 lexical items would have a strong enough connection to concepts to access to their meaning directly. Even though this model stimulated a lot of research on this topic (see De Groot, 1995 and Kroll, van Hell, Tokowicz, & Green, 2010, for review), some authors have questioned the basic principles assumed by the model (see Brysbaert & Duyck, 2010, for a review). Importantly, it falls short at postulating the role of the between-language connection in language production.

Hence, when a concept is activated with the intention of being produced, the activation would spread to both languages and both lexical systems would react if lexical access is not language selective. Crucially, items belonging to the non-target language(s) would also be active and would be potential candidates (and therefore competitors) to the intended word. In contrast to what happens to a monolingual, who needs to produce “dog” and “cat” gets activated and competes for selection, a bilingual who needs to produce “dog” would not only have “dog” and “perro” activated as target words that need to be chosen after language selection, but also, as a consequence of the activation spread to semantic neighbors, “cat” and “gato” (the Spanish for “cat”) would compete for selection as well. Nevertheless, bilingual speakers show great control of this process and rarely fall into language-mixing errors. This implies the existence of mechanisms responsible for language selection. In the case where lexical activation is language non-selective, regulatory mechanisms would prevent the non-target lexical item of the target concept from being the most activated candidate (see below for a description of these regulatory mechanisms). But, do we know whether lexical access is indeed language selective?

Different behavioral paradigms have been used to reveal whether the lexical access in bilinguals is language selective or not. A good example is semantic interference in the picture-word interference paradigm (MacLeod, 1991). This semantic interference (Glaser & Glaser, 1989; Lupker, 1979) is a robust finding that shows larger reaction times when naming pictures with a semantically related distractor word printed on it (naming the picture of a dog with the word “cat” printed on it), as compared to a semantically unrelated distractor printed on it (naming a dog with the word “car” printed on it). These longer reaction times have been argued to reflect competition between candidates of lexical selection (Roelofs, 1992; Schriefers, Meyer, & Levelt, 1990). When several studies tested this same effect but presenting the distractor word in the non-target language (i.e., not the language in which the picture has to be named, see MacLeod, 1991; Smith, 1997, Ehri & Ryan, 1980; Miller & Kroll, 2002), naming latencies were also slower when the distractor in the other language was semantically related (naming a dog with the Spanish word “gato” [“cat”] printed on it) as compared to an unrelated one (naming a dog with the Spanish word “coche” [“car”] printed on it). Between-language effects have been repeatedly shown in the literature: several studies have shown that between-language interference occurs not only at the semantic level, but also at the lexical level (see Marian & Spivey, 2003, for an eye tracking study showing that bilinguals look at phonologically similar distractor objects in either the target or the other language), other studies have shown that syntactic priming exists between languages (Hartsuiker, Pickering, & Veltkamp, 2004) and even cross-linguistic phonetic interference (Flege & Port, 1981). Furthermore, the between- language interference seems to be stronger from the non-present dominant L1 to a non-dominant L2 context (Hermans, Bongaerts, de Bot & Schreuder, 1999). Facilitation also occurs when a concept is repeated both within and between languages (Hernandez & Reyes, 2002) and when a picture to be named is a cognate in both languages (Costa, Caramazza, & Sebastián-Galles, 2000; Hoshino & Kroll, 2008).

Although defenders of each perspective (language-selective and non-language-selective perspectives) have used these behavioral data in their favour by providing different explanations to account for them, recent neuroimaging data shows that the same brain areas are involved in the perception and production of languages that a bilingual speaks (Perani & Abutalebi, 2005). Such studies also show that cortical and subcortical structures use inhibition to solve lexical competition between languages (Abutalebi & Green, 2007; Abutalebi, 2008), and this has been also supported by neurophysiological data (Macizo, Bajo, & Martín, 2010).

Some researchers have recently shown evidence suggesting that certain non-linguistic cues such as interlocutor’s identity can be used for language prediction before any linguistic event occurs (Martin, Molnar, & Carreiras, 2016; Molnar, Ibáñez-Molina, & Carreiras, 2015), thus arguably leading to a language selective access. However, the diverse evidence presented in the previous paragraphs seems to support the notion of language non-selective access, a perspective that is gaining strength from studies showing that both languages that a bilingual speaks are always active, irrespectively of the language in use (Dijkstra, Grainger &

van Heuven, 1999; Sumiya & Healy, 2004; Kaushanskaya & Marian, 2007; Thierry & Wu, 2007). If that is so, control mechanisms that would prevent between language interference are definitely needed. In the next section, I will review the proposed theories that argue for control mechanisms, especially inhibition, that bilinguals use to control the languages that they speak.

III. Language control in bilinguals

It seems clear, in line with the evidence presented above, that "bilingual speakers need to control their production in such a way that the two languages do not end up mixed in an inappropriate manner during the discourse" (Costa & Santesteban, 2006; p.115). Many attempts have been made to postulate different controlling mechanisms that would operate to guarantee lexical selection in the intended language, and not in the non-target language, preventing interference (Green, 1998; Hermans et al., 1998; Lee & Williams, 2001; Costa & Santesteban, 2004a; b).

To try to account for this mechanism, as mentioned in the previous section, a sort of *mental switch* was originally proposed, which would turn on and off each language, allowing the speaker to use one language efficiently without the intrusion of the one that is "off" (Penfield & Roberts, 1959; McNamara & Kushnir, 1971). This idea was soon discarded because of various theoretical and empirical criticisms (Paradis, 1981). Later, it was assumed that different languages can have different levels of activation, and that in order to efficiently choose the intended language, its activation level must exceed that of the non-target language (Grosjean, 1988; 1997; 1998). Different authors argued in favour of control mechanisms that would allow bilinguals to be in different language modes: sometimes they would need to speak one only language and to exclude the others, while in other cases mixing might be allowed (Grosjean, 1985; 1997)

One of the most (if not the most) convincing proposals concerning the issue of language control in bilinguals was proposed by Green in 1998. This seminal article introduced *the inhibitory control* (IC) model, which argues that both languages that a bilingual speaks compete with each other to control the output and that inhibition should be applied to the unwanted candidates. The model is grounded on two assumptions: firstly, it assumes that the regulation of languages and control of action (see below) have much in common (Macnamara, Krauthammer, & Bolgar, 1968; Paradis, 1980). Secondly, it states that the regulation of languages is achieved by the modification of levels of activation within and between language networks (as opposed to a "switch device", De Bot & Schreuder, 1993; Grainger & Dijkstra, 1992; Grosjean, 1988).

For the first assumption, which states that the regulation of languages is equated to control of actions, the classic models of behavior control (e.g., Norman & Shallice; 1986) are a clear way to understand the processes involved in language control. There are various different actions that we manage to perform successfully, and they can range from completely routine to completely new. To achieve success in each specific task, the existence of *mental schemas* is proposed, some sort of mental compendium of actions that is triggered to achieve a set goal. For example, deciding to make a coffee would activate different schemas, such as filling a kettle or preparing the cups, each involving processes that would eventually activate other sub-schemas until the main task is completed. When the task has been previously conducted, automatic performance (e.g. driving) or semi-automatic performance (e.g. dressing) can be achieved, whereas in situations in which the task is new or automatic control is not sufficient, the process must be administered by a supervisory attentional system (SAS). This system would construct or modify existing schemas and monitor their performance oriented to the goal achievement (Shallice, Burgess, & Robertson 1996).

When Green applies these same processes to language, the system is equivalent: as can be seen in Fig. 5, the IC model postulates that the communicative language process is a product of the interaction between the SAS (as explained above in control of actions), the language schemas (the equivalent to action schemas, mechanisms that would be in charge of the correct use of a specific language or of a part of it, which makes sure that all its rules are followed) and lexico-semantic systems of the languages. Firstly, there is a main goal to be achieved, which is the main information to be communicated. It drives the conceptualizer to build the concept to be transmitted. The conceptualizer (which is language independent) transmits the conceptual information to the lexico-semantic system, and the lexical candidates (often called "lemmas") are activated. The SAS specifies the language to be used to the language task schema system, and then a language schema is retrieved, applied to the lexico-semantic system and used. This language schema will regulate the outputs of the lexico-semantic system by activating the desired outputs (e.g., the lemmas) and inhibiting any other competing information. This schema will be active until the goal is achieved, another schema inhibits the current one, or the SAS changes the goal. When different schemas compete with each other for activation, interference arises. For example, within language schema-competition can easily be seen in the classic Stroop (Stroop, 1935) effect, which can be elicited by competition between "colour naming" schemas and "word reading" schemas, a competition that of course would be stronger in the incongruent condition where the colour name and the printed word are different.

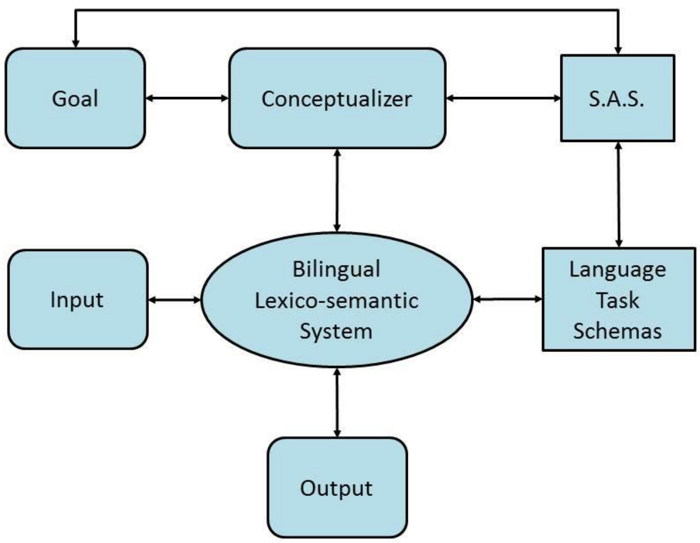

Figure 5. Explanatory sketch of the inhibitory control model. Adapted from Green (1998).

When the selection of language needs to be managed by means of these schemas (i.e., between language schema-competition), things get more complicated. As explained above, the conceptualizer transmits the concept to be produced, which is language-independent. As previously proposed by Levelt (1989, Levelt, Roelofs, & Meyer, 1999), the concept activates its corresponding lexical items (or lemmas) and its syntactic properties. The activation of a concept activates the lemmas of both languages (which have different language tags, i.e., L1 or L2), and the IC model proposes that the lemmas with the incorrect language tags have to be suppressed by the SAS reactively, via inhibition. Thus, this model postulates that SAS reflects the intention of performing in a certain language (L2 or L1) by activating different language schemas that compete for output control. Different lemmas would be activated or suppressed based on their language tags, leading to the final selection of the lexical items corresponding to the leading language schema. The inhibition is reactive, so the (re)activation of specific lemmas would require input from the outside or the conceptual level. It takes some time to return to the pre-inhibition state if, for example, a language that has been the non-target one in a concrete situation needs to be used again. Empirical evidence of this can be observed in switching tasks in which bilingual participants have to name objects in both of their languages. Because of the inhibition of the non-target language, naming in one language right after speaking in the other language (switch) is harder and takes longer than when the utterances are produced in the same language without switching (see, among many others, Thomas & Allport, 2000). Furthermore, and similarly to what happens in non-linguistic physical actions, some schemas are more automatized than others. When a non-automatized schema (e.g., L2) is activated and the competing schemas are much stronger and automatized (e.g., L1), the inhibition applied to the competing schema is proportionally increased. In line with the prediction that the stronger competitors should be strongly inhibited, it has been repeatedly shown that it takes longer to come back to the stronger and more dominant L1 after naming in L2 (where the stronger L1 had to be very intensely inhibited) than the other way round (Meuter & Allport, 1999).

Other theories have been proposed to account for bilingual language control in production (see, for example, Green & Abutalebi's *adaptive control hypothesis*, 2013, which takes into consideration the demands of the context and how they change the adaptive mechanisms), and also in perception, and they all mainly rely on inhibition. In the domain of perception, *the bilingual interactive activation* model (BIA, see Dijkstra, Van Heuven, & Grainger, 1998) assumes that language perception is initially language non-selective, and has different stages of processing that happen when a written word is encountered: the sublexical orthography (which spreads its activation to sublexical phonology), lexical orthography (which spreads to lexical phonology), language nodes and semantics. Potential candidates within and between languages are activated due to the spread activation of the perceived features, and language nodes control this competition by identifying the target language that the individual is working on, and inhibiting the potential candidates that are active in the non-target language.

The IC model provides a grounding for the differences between bilinguals and monolinguals and their consequences (Abutalebi & Green, 2008; for a neurocognitive adaptation of the IC model) and therefore I will assume, for the purposes of this thesis, that language competition in production between bilinguals' L1 and L2 is solved mainly applying inhibition to the non-target language, as proposed by the IC model (but see Costa, Miozzo, & Caramazza, 1999; Costa & Caramazza, 1999; for a different view).

So far, I have explored the current perspectives on language organization and access in both monolinguals and bilinguals. The difference between the two groups is twofold, quantitative and qualitative. The first one is purely linguistic, due to the fact that they speak a different number of languages. The impact of this difference is worth exploring: Does this have any linguistic consequence? Does a bilingual differ, in each of the languages, from a monolingual who speaks only one of them? These consequences in the linguistic level are going to be discussed in the next section. The second difference stems not from the amount of languages that bilinguals speak, but from the status and the quality of those two languages. The two languages seem to affect each other and to send each other activating and inhibiting signals, and it seems reasonable to assume that this interconnection might vary depending on the proficiency in each of the languages (Kroll & Stewart, 1994). Thus, bilinguals need to manage their two languages via different control mechanisms, which suggests that some non-linguistic and high-level general cognitive skills might have different importance in a bilingual compared to a monolingual brain. As suggested by the IC model, bilinguals require much stronger control mechanisms than monolinguals, and they face heavy demands that require control and inhibition (and not only activation) to simply use one language and not the other. Does this have any other impact at some cognitive level? If these high level abilities are general, and thus not domain specific (see Fodor, 1983), they would be involved in any action that would require the use of inhibitory control for both bilinguals and monolinguals. If bilinguals, as a result of managing two languages, make more and better use of this control mechanism, it could be enhanced and this enhancement reflected in non-linguistic tasks that require those same control mechanisms. If it is a function specifically involved in language, it would be used by bilinguals only but it would only affect their linguistic skills. In the next section, I will explore the differences between these two groups in both linguistic and non-linguistic features.

IV. The linguistic consequences of bilingualism.

There is a general consensus on the linguistic consequences of being a bilingual, and while some of those consequences speak of benefits, others indicate that bilinguals might suffer from some disadvantages. However, it is worth noting that these consequences do not imply any kind of limitation in real life situations and are mainly reported under experimental conditions.

1. The negative linguistic impact

Firstly, when we look at the negative consequences, the results reported in the literature appear to be consistent. One of the most crucial differences between bilinguals and monolinguals that has been consistently reported is that bilinguals show a poorer knowledge base in each of the languages, which affects all the language processes built on this base (Bialystok, Craik, Green, & Gollan, 2009; Bialystok & Luk, 2012). Hence, it has been repeatedly found that bilinguals generally know fewer words, i.e. *a smaller vocabulary*, in each of their languages than monolinguals (Verhallen & Schoonen, 1993; Vermeer, 1992; Perani et al., 2003; Portocarrero, Burright & Donovick, 2007). This vocabulary gap is consistent during childhood (Mahon & Crutchley, 2006; Oller & Eilers, 2002, Bialystok, Luk, Peets, & Yang, 2010), and it is of crucial importance for bilingual children because the vocabulary sizes are often taken as a measure of children's linguistic, cognitive and academic development (Ouellette, 2006; Ricketts, Nation, & Bishop, 2007; Swanson, Rosston, Gerber, & Solari, 2008). When inferences related to development are measured through language-based tests performed in children's second language, they might inappropriately indicate a delay in bilinguals compared to monolinguals, who were tested in their only language. However, well-controlled studies have shown no cognitive differences between monolingual and bilingual children (Baker & Jones, 1998; Cook, 1997; Hakuta, 1986). Moreover, even though the bilinguals' acquisition was at first assumed to be delayed compared to that of monolinguals (Oller, Eilers, Urbano, & Cobo-Lewis, 1997; Pearson, Fernandez & Oller, 1993), it has been recently shown that bilinguals and monolinguals achieve their language milestones in each language at the same time as monolinguals (Petitto et al., 2001; Petitto & Kovelman, 2003; Petitto & Holowka, 2002) and it has been argued that the semantic and conceptual development is the same (Holowka, Brosseau-Lapré, & Petitto, 2002).

Additionally, despite the enormous variation in vocabulary size in adults, we tend to observe the same lexical gap between adult bilinguals and monolinguals (Bialystok, Craik, & Luk, 2008; Portocarrero, Burright & Donovick, 2007). Given the variability in adults' knowledge, which makes the measurement of vocabulary size in this age group difficult, these differences have been explored by studying *lexical access* in two different processes: lexical retrieval (the speed with which target words can be retrieved and produced) and verbal fluency (the number of words that can be produced according to a specific criterion, e.g. "tell me all the words you can that start with the letter "P" or "tell me all the animals you can"). There is an extensive body of evidence showing differences in those two aspects. When lexical retrieval abilities of bilingual speakers are under study, the most commonly used tasks are those that require participants to name pictures in both languages (Costa & Santesteban, 2004), to make semantic classifications for words in both languages (Dufour & Kroll, 1995), or to translate words from one language to another (Kroll & Stewart, 1994). Bilinguals' performance on these tasks is compared between the languages or to the performance of monolinguals. One of the most used tasks for this purpose is the Boston Naming Task (BNT, see Kaplan, Goodglass, & Weintraub, 1983), where participants have to name the picture that they are presented with. When bilinguals and monolinguals are tested in the BNT, bilinguals produce less correct responses (Roberts, Garcia, Desrochers, & Hernandez, 2002; Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007) and make more errors when the task is speeded (i.e., the maximum time to produce each trial is limited, Bialystok et al., 2008). When the task is timed (the time required to produce the answers is measured), bilinguals usually perform slower than monolinguals (Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Similarly, bilinguals report having more tip of the tongue experiences (Gollan & Acenas, 2004) and, when recognizing words, bilinguals show poorer word identification in a noisy context (Rogers et al., 2006), and more interference in lexical decision tasks (Ransdell & Fischler, 1987). These deficits in lexical access are not restricted to infancy and adulthood, but they persist with aging (Gollan et al., 2007). All these results seem to indicate that the act of retrieving a lexical item is more effortful for bilinguals than for monolinguals, either because they have less items available or because the access to them is harder.

It is worth noting that the disadvantages in bilinguals' lexical access described above are sometimes under debate. When bilinguals and monolinguals are tested in these paradigms, the language used is sometimes the bilinguals' weaker language, L2 (Roberts, Garcia, Desrochers, & Hernandez, 2002). This makes the comparison unfair, given that bilinguals do not usually have the same proficiency in both languages and, even though their language knowledge allows them to live and communicate normally, their skills in that language might still be inferior to those of monolingual speakers. Consequently, many authors have explored the situations the possible bilingual disadvantages appear in and those in which they do not. Gollan and her colleagues (2002) found that bilinguals were especially impaired in semantic fluency tasks (as compared to letter and proper name fluency tasks) and, although they consistently produced fewer exemplars in all the tasks, both monolinguals and bilinguals produced the same amount of errors. Gollan, Montoya and Bonanni (2005) also show similar tip-of-the-tongue effects for both language groups when naming proper names. In contrast, bilinguals show more tip-of-the-tongue effects in easy target words but less in hard target words compared to monolinguals (Gollan & Brown, 2006). Surprisingly, Gollan and colleagues (2005) reported that, even though bilinguals name pictures in their dominant language slower and more erroneously than monolinguals, the naming times did not differ between groups when participants named the pictures several times or when they had to classify them as human made or natural. Additionally, bilinguals named the pictures faster than monolinguals if they knew the name of the picture in both of their languages. The authors concluded that the results obtained indicate that the difference between bilinguals and monolinguals can be situated after conceptual processing and that the implicit lexical activation in the non-target language can facilitate retrieval in the target language.

When verbal fluency is looked at, the classical measurement is to ask participants to name as many candidates as they can that belong either to one category (e.g., animals, semantic fluency) or that start with a concrete letter/sound (phonological fluency) in a limited amount of time, that is usually 60 seconds. These two processes are different in their approach: while category fluency somehow resembles the natural, everyday word retrieval procedures during language use (where the semantic context primes and facilitates the production of the upcoming words by association), phonological fluency is not present in daily life conversations (that usually do not require the production of words that have the same first letter). This difficulty, together with extra restrictions that the task imposes (repetitions of words in different forms are not allowed), means that the letter fluency task requires greater involvement of some executive functions such as monitoring and working memory. Namely, while category fluency is a more direct indicator of the vocabulary size of a concrete field (e.g., how many animals you know), letter fluency imposes additional demands to keep track of the words already produced and to look for more candidates that satisfy the stipulated criterion. This distinction is reflected in brain region activation, with bilinguals showing activation in different areas for each of the tasks (see Grogan, Green, Ali, Crinion, &

Price, 2009). Typically, monolinguals outperform bilinguals in both tasks, but the differences are greater in the category fluency task (Bialystok et al., 2008, Gollan, Montoya, & Werner, 2002; Portocarrero et al., 2007). Somewhat surprisingly, college students who were studying in a country with a different language to their own produced fewer words in a verbal fluency task when compared with peers who had not gone abroad to study (Linck, Kroll, & Sunderman, 2009), but this difference disappeared after returning home. However, it is worth mentioning that the longitudinal evidence reported by Woumans, Surmont, Struys, and Duyck (in press) show a different pattern. They tested children who were initiating either a bilingual or a monolingual kindergarten, and again one year later. Among other cognitive changes (see the subsection "Cognitive Disadvantages"), the authors reported similar verbal fluency of both groups, both at baseline and one year later.

Different explanations have been proposed to account for the bilinguals' generally lower performance in tasks that tap into verbal fluency. As a first option, the difference could be ascribed to the abovementioned difference in vocabulary size. Rosselli et al. (2000) argue that, if bilinguals indeed have less lexical items than monolinguals, this would be reflected mostly in the category fluency task and would not be so evident in the letter fluency. Bilinguals would know fewer number of exemplars from each category (for example, bilinguals would know less exemplars of "tree names" than a monolingual in the shared language) but the amount of words known that start with the same letter would be big enough to show no relevant differences. The authors tested this hypothesis and found differences in the category fluency but no differences in the letter fluency task (see Rosselli et al., 2000).

An alternative explanation is based on the hypothesis that the existing competition between the languages (see again the model proposed by Green, 1998) is what causes the difference, hampering the performance in verbal fluency tasks. Complementarily, given that bilinguals use each of the languages less often than monolinguals, the connections between words that are required to perform the production tasks are weaker than those of monolinguals (Michael & Gollan, 2005; Gollan, Montoya, Cera & Sandoval, 2008). This perspective is supported by the results reported in the very beginning of this subsection from picture naming tasks where bilinguals take longer to retrieve each item, which might indicate that the time limit of the fluency tasks could be limiting bilinguals' performance. As stated previously, bilinguals might need to deal with some competition arising from the strong activation of both languages, and dealing with that competition might take some time and delay their answers. Alternatively, they might just need more time to reach each lexical entry due to weaker connections in the system. In that case, bilinguals would not have smaller vocabulary sizes, but they would just need more time to retrieve all the possible candidates. In category fluency tasks, competing elements (for example, animal items of both languages) would have a much more important role than in letter fluency, which is not so semantically charged and thus competition due to activation of the exemplars in the other language would be weaker. In order to disentangle this question and obtain a clearer picture of whether the effects are due to vocabulary size or to competition-caused delay, several studies conducted verbal fluency tasks and analyzed how the words were produced in each minute allotted to each trial. If the difference is due to vocabulary size, bilinguals should produce most of their words in the beginning of the minute and very few in the end, reflecting that the pool of words has been almost completely used, and monolinguals should keep on producing words later than bilinguals. On the contrary, if the bilinguals' lower performance is a matter of more time needed due to competition, bilinguals should produce each item slower and they should keep on producing words later than monolinguals, indicating that their mental lexicon was not yet used up but that they just need more time to select and retrieve each of the exemplars. Two studies (Sandoval, Gollan, Ferreira, & Salmon, 2010 and Luo, Luk, & Bialystok, 2010) analyzed the time course of the word production in fluency tasks, and showed that bilinguals did not stop producing earlier than monolinguals, but they kept on producing words later in the assigned minute. The authors ascribe the results to the interference of the non-target language (or weaker links in the system, given that the two explanations are not mutually exclusive), rather than a restriction imposed by a limited vocabulary size. Interestingly, when Luo and colleagues (2010) matched monolinguals and bilinguals in vocabulary, both groups performed similarly in category fluency task and bilinguals outperformed monolinguals on the letter fluency task. This led the authors to conclude that bilinguals benefited from a better use of working memory and monitoring abilities. Moreover, when vocabulary is taken into account and monolinguals, bilinguals with matched vocabulary to those of monolinguals, and bilinguals with lower vocabulary level are tested, bilinguals with matched vocabulary size outperformed their monolingual peers (Bialystok, Craik, & Luk, 2008). These results bring to light the necessity of matching groups in relevant factors such as vocabulary size.

Other evidence supports the notion that competition between languages is what slows down lexical access: when a picture has to be named while an audibly played word in the other language has to be ignored, naming latencies are longer if the word to be ignored is phonologically related to the actual utterance (Hermans, Bongaerts, de Bot, & Schreuder, 1998). Similarly, when participants were presented with two pictures which were colour-coded to indicate one language or the other and were asked to name one of them while ignoring the other, participants showed longer interference (i.e., longer reaction times) if the distractor picture's label was a cognate word (Colomé, & Miozzo, 2010).

To overcome these interfering effects, bilingual speakers seem to rely heavily on social visual and familiarity cues that they attach to a language. Thus, naming in Chinese was facilitated if the picture to be named was preceded by an Asian face as compared to a Caucasian face (Li, Yang, Scherf, & Li, 2013), but Chinese faces also decreased Chinese immigrants' English fluency when they had to talk to them instead of to a Caucasian face (Zhang, Morris, Cheng, & Yap, 2013). Molnar, Ibañez-Molina, and Carreiras (2015) showed that previous experience also counts, given that the comprehension of words produced in the language that a face was previously associated with was faster than when language and face were not previously associated (see also Woumans et al., 2015 for strong face-language associations that get weaker if the face is perceived as bilingual).

2. The linguistic benefits

As the previous section made clear, the negative linguistic consequences of bilingualism are not such major disadvantages, and the specific situations in which these effects appear and the reasons behind these phenomena are still under debate. Some potential effects of bilingualism on other linguistic aspects, such as phonologic awareness, are still inconclusive (Verhoeven, 2007; Marinova-Todd, Zhao, & Berhardt, 2010; Laurent & Martinot, 2010, and Bialystok, Majumder, & Martin, 2003). However, bilingualism also provides bilinguals with some advantages in linguistic skills. The first and probably the most important one, is the ability to communicate and perceive the world in *two languages*.

Additionally, bilingualism does have positive effects on *word and language learning*: Kaushanskaya and Marian (2009) tested two different bilingual groups (English-Spanish and English-Mandarin) and a monolingual group in a word learning task in their native languages, and showed that both bilingual groups outperformed monolinguals. Bradley, King and Hernandez (2013) tested English monolinguals and early sequential Spanish-English bilinguals in novel German word learning, and found that, even if they did not differ in behavioral performance, bilinguals made a more efficient use of their brain networks. Related to this, it is not surprising that bilinguals have been shown to outperform monolinguals when a third language has to be learnt (Cenoz & Valencia, 1994; Swain, Lapkin, Rowen, & Hart, 1990; Sanz, 2000).

The origin of this better performance in new language learning might be due to the *flexibility* that bilingual infants show when compared to monolingual peers in word learning: monolinguals are usually guided by the "mutual exclusivity" principle, according to which infants cannot acquire a new label for an already learnt and labeled object, and therefore they attribute it to a new unlabeled object or to a salient feature of the known object (Markman & Wachtel, 1998). In contrast, bilingual children are not restricted by this, and they can easily apply a newly presented label to an object that they already knew the name of, reflecting their ability to have different names in different languages for the same object (Kandhadai, Hall, & Werker, 2016). As a further evidence for the flexibility of the bilingual children, preverbal bilinguals have been shown to simultaneously learn multiple speech structures better than monolinguals (see Kovács & Mehler, 2009; although this finding has never been replicated).

Furthermore, despite similar patterns of distinction of different spoken languages shown by bilingual and monolingual children (Byers-Heinlein, Burns, & Werker, 2010; Bosch & Sebastián-Gallés, 2001), bilingual children also show better *language sensitivity* based on visual cues. Monolingual children are able to distinguish silent video recordings of sentences in their native language from another language at 4 and 6 months of age, but they lose the ability by 8 months of age. However, 8-month-old bilinguals could distinguish not only silent video recordings of familiar languages (Weikum et al., 2007), but also of unfamiliar languages (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012).

The linguistic consequences of bilingualism for language skills are important but the impact may go beyond the linguistic domain. Given the purpose of this thesis, in the next section I will focus on the cognitive consequences that bilinguals might experience due to the extra abilities that bilinguals seem to need to control and efficiently manage their languages (Green, 1998).

V. *The cognitive consequences of bilingualism.*

It has been repeatedly shown in the literature that various life experiences can have a significant impact in human cognition, at both behavioral and neurological levels. For example, it has been shown that video gaming experience enhances visual selective attention (Green & Bavelier, 2003) and that, with training, that ability can increase further (Feng, Spence, & Pratt, 2007). Architects, as compared to non-architects, show better visuo-spatial abilities (Salthouse & Mitchell, 1990). Structural brain changes have been reported in London taxi drivers (Maguire et al., 2000). They showed enlarged regions of the hippocampus responsible for spatial navigation as a consequence of their daily route-finding experiences. Similarly, musicians whose instruments depend on the use of the fingers of the left hand have shown an increased cortical representation of those important fingers (Elbert et al., 1995). Importantly, different lifestyle experiences have a direct impact on crucial aspects of individuals' quality of life: for instance, an active and socially integrated lifestyle with intellectual challenges (such as education and occupation) accumulate the cognitive reserve and can protect against neurodegenerative processes such as dementia or Alzheimer's disease (AD) and even delay them (Fratiglioni, Paillard-Borg, & Winblad, 2004; Staff, Murray, Deary, & Whalley, 2004).

If experiences and abilities that are generally acquired late in life and are used few hours per day can have such a huge impact on domain general cognitive abilities, one could tentatively suggest that the knowledge and use of two languages as compared to a single one should have an impact as well. Many bilinguals around the world acquire both languages either at home or at school when they are young, and use them both on a daily or regular basis. Arguably, the ability to use two different representations of one of the most complex human abilities (i.e., language) should create important changes at different levels of cognition. To begin with, when the brains of bilingual and monolingual speakers are compared, people who speak two languages show an increased grey matter density in the left inferior parietal cortex and, importantly, this difference is more accentuated in early bilinguals and highly proficient bilinguals (Mechelli et al., 2004). It is not surprising that bilingualism, which can shape the speaker's brain, can also produce cognitive changes in domain-general abilities. As we shall see in the upcoming subsection 3, many researchers have argued that there can be some kind of transfer on the training that bilingualism provides to general abilities not related to language. But the potential effects of bilingualism on human cognition are many, and many are the fields that researchers have explored trying to capture an index of the impact. This has led to some results for which there is general agreement, and to others that are more debatable. More generally, some results speak of an advantage while others show a disadvantage when bilinguals' abilities are compared to those of monolinguals (Bialystok, 2009).

1. The cognitive disadvantages

When it comes to the cognitive disadvantages of bilingualism, there is very little evidence of drawbacks, and those that do show up are present in very few aspects of cognition. It is worth mentioning that, *historically*, bilingualism has been considered to have a negative impact on some cognitive abilities, such as IQ, since the topic first caught researchers' attention almost one century ago (see Darcy, 1946, for a review of evidence from that time). Thus, Smith (1923) found that bilingual children showed poorer performance in vocabulary, expression and "accuracy of thought" abilities when compared to their monolingual peers. Saer (1923) tested 1400 children in Wales, reported a significant advantage of monolinguals over bilinguals in IQ only in rural children, with no significant differences in urban children, and concluded that bilingualism created "mental confusion". Similarly, Goodenough (1926) named this disadvantage a "mental retardation". Jones and Stewart (1951) found that bilingual children performed poorly in verbal IQ tests when compared to monolinguals. However, a few decades later some researchers started pointing out that previous findings showing bilingual disadvantages were a consequence of uncontrolled external socio-economic factors (James, 1960). As an illustration, Lewis (1959) published data arguing that monolinguals showed higher nonverbal IQ scores as compared to bilinguals, but Jones (1960) criticized the Lewis study for failing to control for socioeconomic status. These same criticisms could be applied to many studies showing a monolingual advantage in IQ without controlling for socio-economic status (Graham, 1925; Mead, 1927; Rigg, 1928; Wang, 1926; see also Darcy, 1953 for a review of the methodological concerns on the findings), because bilinguals of those studies often had a lower SES (McCarthy, 1946). Some years after, the field seemed to agree that bilingualism might be associated with different social variables that could produce bilinguals' lower performance in IQ (especially in the non-verbal IQ) such as socioeconomic status, and that there was no consistent data to indicate that bilinguals systematically score lower in IQ tests (Darcy, 1963). Eventually, Peal and Lambert (1962) reported that, when bilingual and monolingual samples are properly matched in relevant factors (i.e., SES, age, sex), bilinguals surprisingly outperformed monolinguals in verbal and especially non-verbal tasks, showing what they called a greater "mental flexibility" and giving birth to the idea that would later be the *"bilingual advantage"* (see below in the subsection 3, "The bilingual advantage hypothesis"). This was later confirmed (Ben-Zeev, 1977) and recently explored and confirmed again by Woumans et al. (in press), who showed that children who went for a year to a bilingual kindergarden outperformed peers that went for a year to a monolingual kindergarten on intelligence, but not on verbal fluency or cognitive control.

Similarly, a common belief was that bilingualism could be more effortful than acquiring just one language (Macnamara, 1967; Torrance, Wu, Gowan, & Aliotti, 1970), and thus potentially create a language acquisition delay or confusion when comparing bilinguals to monolinguals. However, even if some small delays have been reported (Byers-Heinlein, Burns, & Werker, 2010), it is important to note that bilingual children have been shown to reach their language development milestones at the same time as monolinguals (Oller, Eilers, Urbano, & Cobo-Lewis, 1997; Petitto, Katerelos, Levy, Gauna, Tétreault, & Ferraro, 2001).

As can be seen, the original fears and disadvantages have been discredited by research, and very few disadvantages are still discussed or accepted nowadays in the scientific domain. In the general cognitive field, recently Folke and colleagues (2016) wanted to test differences in *metacognition* caused by bilingualism. Metacognition is a construct that defines the ability to evaluate one's own performance (Fleming, Ryu, Golfinos, & Blackmon, 2014), and it is usually defined in two steps – the action taken and the subjective evaluation of that performance (Nelson & Narens, 1994; Grimaldi, Lau & Basso, 2015). Folke and colleagues' (2016) rationale for this study was straightforward: the languages that a bilingual speaks are always active (Van Hell & Dijkstra, 2002) and the current theories argue for executive functions (Green, 1986, 1998; Green & Abutalebi, 2013) to control language activation. Given that an enhancement of executive functions in bilinguals has been suggested (Bialystok, Craik, Klein, & Viswanathan, 2004; see the section, "the bilingual advantage hypothesis", for a more detailed description of this field), and that a correlation between them and metacognition is consistently found (Fernandez-Duque et al., 2000; Souchay & Isingrini, 2004; Del Missier, Mantyla, & Bruine de Bruin, 2010), it could be expected that metacognition skills are enhanced in bilinguals. In order to explore that possibility, Folke et al. (2016) asked bilingual and monolingual participants to judge which of two circles contained more dots. Subsequently, they had to say how confident they were about the decision they had made. Results indicated that monolinguals and bilinguals did not differ in accuracy but, surprisingly, they found a bilingual disadvantage in metacognitive processing, indicating that monolinguals were better (around 10%) at assessing their own performance. They conclude that there is a possibility that the relation between metacognition and executive functions are different when measured in different domains (Fleming et al., 2014), making the link between them domain-specific.

Hence, it can be seen that only relatively small cognitive disadvantages have been found in experimental situations. As it was explained in this section, the majority of the reported disadvantages, which are argued to be the consequences of bilingualism, were rather unfounded and probably consequence of uncontrolled factors.

2. The cognitive benefits

Cognitive advantages of bilingualism have been reported in the last several years by researchers of different fields, some of which are widely accepted, others much less so. Among the uncontroversial findings, bilinguals have *better appreciation of other people's beliefs* and better use of this knowledge. Take, for example, the following false-belief situation used with children: two puppets are playing with a toy and, when they finish, one of them puts the toy in a box and leaves. The other puppet moves the toy to a different box while the first puppet is away. When the first one comes back, the child (who was watching the entire scene) is asked about where the first puppet will look for the toy. Children 4 years old typically answer correctly "in the original container where she left it", but younger children respond "in the container where the toy is now". However, bilingual children reach the point at which they can understand what the puppet believes earlier in life, and answer correctly at around 3 years of age (Goetz, 2003; Kovács, 2009). It has been argued that this could be derived from bilingual children's enhanced sociolinguistic sensitivity to the language use of interlocutors (Genesse, Boivin, & Nicoladis, 1996). A similar task has been used to replicate the findings of benefits of bilingualism in overcoming false egocentric beliefs (Rubio-Fernández & Glucksberg; 2012).

Following the same rationale, Fan, Liberman, Keysar and Kinzler (2015) tested bilinguals, children exposed to two languages (but not bilingual) and monolinguals in a task that required *taking the perspective of the speaker.* As explained in the previous paragraph, bilingual children have been shown to outperform monolinguals in theory-of-mind tasks that require dealing with false-belief situations (Kovács, 2009; Rubio-Fernández, & Glucksberg, 2012). Additionally, it is known that bilingual children outperform the monolinguals in a *mental rotation* task, which requires taking spatial perspective (Greenberg, Bellana, & Bialystok, 2013). These two findings together seemed to indicate that bilingualism could have an impact on taking the interlocutor's perspective. Fan and colleagues (2015) tested bilinguals, monolinguals and children exposed to more than one language in a task that required identifying the target object by understanding descriptions from the interlocutor's perspective. This task required participants to be able to put themselves in the perspective of their interlocutor and to mentally rotate the observed scenario to understand what the interlocutor sees: some items that were relevant for the task could be seen only by the participant, and others by both the participant and the interlocutor. The participants had to understand that the items that their interlocutors were referring to during the instruction could only be the ones that both the participant and the interlocutor were able to see. The results of the study showed that the bilinguals and children exposed to more than one language identified the target significantly more often than monolinguals, with no differences between the first two groups. They argue that exposure to multiple languages (and not necessarily bilingualism per se) requires being aware of other people's linguistic perspective (Yow & Markman, 2011).

A different perspective on exploring benefits derived from bilingualism comes from the study by Stocco and Prat (2014). The authors maintain that bilinguals are better than monolinguals when there is a need to select and apply *new rules*, because they do this more often than monolinguals do. For example, when they talk to multiple speakers or when they have to rapidly choose and combine linguistic rules according to the language in use, bilinguals need to choose one specific language and then be aware of its particular set of rules in pragmatics, phonetics and syntax. Thus, when a Basque-Spanish bilingual switches from having a conversation in Basque to talking to a Spanish speaker, she must be aware that the rules have changed: for instance, the Basque way of pluralizing a noun, adding "-ak" at the end of it, is no longer valid, and she should add an "-s" instead. The authors tested this hypothesis by comparing monolinguals and bilinguals in Rapid Instructed Task Learning (RITL), in which the tasks are generated by either combining tasks in which participants have to give a different response to different kinds of stimuli (Hartstra, Kühn, Verguts, & Brass, 2011) or by combining different arithmetic operations (Stocco, Lebiere, O'Reilly, & Anderson, 2012). Given that bilingualism imposes the need to change from one set of rules to another and combine the rules quickly to efficiently reach the ultimate goal of communication, Stocco and Prat (2014) predicted that the bilinguals should perform better on the RITL. To test that hypothesis, they trained monolingual and bilingual participants in sets of tasks within the RITL paradigm. In a second session of the experiment, on a second day, the participants were provided with the tasks that they had already performed the previous day, together with new tasks that they had not encountered before. The results showed that bilinguals and monolinguals performed equally on the previously encountered tasks, but that bilinguals were faster than monolinguals on the new tasks. This confirmed their hypothesis of bilinguals being better than monolinguals when there is a need to adapt to a new set of rules.

Lastly, bilingualism has been associated with an *improvement in executive functions*. That improvement, which has been called "the bilingual advantage" in the literature (Bialystok, 1999; Bialystok & Martin, 2004; Bialystok, 2011), is taken to show that bilinguals are better than monolinguals in executive control and, in particular, inhibition. The reasoning behind this is that bilinguals have to constantly deal with their two languages and inhibit the non-target one (see again the IC model, Green, 1998, in the section III of this chapter), which serves as training in those abilities and is eventually transferred to a general enhancement (not restricted to language situations). This enhancement, the so-called bilingual advantage, should be easily observed in any situation in which the use of executive functions (and in particular inhibition) is needed.

As the relation between bilingualism and enhancement in executive functions is in the focus of this thesis, it will be further discussed in the following sections.

3. The bilingual advantage hypothesis

The "bilingual advantage" has drawn much research attention (see Bialystok, Craik, Klein, & Viswanathan; 2004) and has been widely debated (Paap, Johnson & Sawi, 2015a; Duñabeitia & Carreiras, 2015).

The concept of executive functions (EF), as proposed by Miyake & Friedman (2012, see also Miyake et al., 2000) is generally described as a construct that encompasses inhibition (i.e., the ability to suppress dominant or salient responses), shifting (the capacity to switch between tasks or mental sets), and updating (the ability to constantly monitor and to rapidly update the information in the working memory). It is one of the most important general cognitive skills that human beings have, and is present in almost any situation of our daily lives (Mischel et al., 2011; Moffitt et al., 2011), mainly because our self-control or willpower depend on it. These skills have a large impact on socially important behaviors (Friedman et al., 2007; 2011; Young et al, 2009). Importantly, although genetics play an important role in the EF skills of individuals (Friedman et al., 2008), they can be improved by training (Moreno et al., 2011; Karbach & Kray, 2009; Kray & Lindenberger, 2000; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008).

The proponents of the bilingual advantage hypothesis suggest that bilingualism could be considered extra executive function training that monolinguals do not undergo, because bilinguals use their executive functions much more often than monolinguals. Considering all the evidence presented supporting this hypothesis, this assumption seems reasonable. Furthermore, research that examined many different populations and used different techniques and paradigms has repeatedly shown that both of the languages that a bilingual speaks are always active and available, even if only one is present in the current context (see, among others, Hernandez, Bates & Avila, 1996; Dijkstra, Grainger, & van Heuven, 1999; Sumiya & Healy, 2004; Kaushanskaya & Marian, 2007; Thierry & Wu, 2007). This evidence has been reported in both behavioral (Beauvillan & Grainger, 1987; Colomé, 2001; Grainger, 1993; Hernandez, Bates, & Avila, 1996; Kroll & de Groot, 1997) and neuroimaging studies (Marian, Spivey, & Hirsch, 2003; Martin, Dering, Thomas, & Thierry, 2009). Additionally, there is a consensus that language selection is part of bilingual speech production and does not occur prior to it (Kroll, Bobb, & Wodniecka, 2006). This makes the demands that monolinguals and bilinguals face very different. It is true, however, that both face situations in which they have to select between competing within-language candidates (for example, close semantic neighbors such as cup and mug in English, Luce & Large, 2001), so both monolinguals and bilinguals would be equally used to applying lateral inhibition to suppress strong within-language candidates. In addition to the within-language competition, bilinguals have to deal with between-language competition too, stemming from the candidates from different languages for the same concept (following the example before, “cup” and “taza”). This situation is, obviously, not faced by monolinguals, and it creates an obstacle that needs to be solved: competition between the two languages. Bilinguals need to choose the language they want to use, pick the lexical form of the item to be used and assemble it in a logic and coherent output that follows the rules of the chosen language. Additionally, they have to do so while avoiding any extraneous influence from the non-target language. As explained in the “language control in bilinguals” section (section III of Introduction), one of the most extended views suggests that the conflict between the two languages has to be monitored and the non-target language has to be inhibited (Green, 1998; Levy, McVeigh, Marful, & Anderson, 2007) by mental schemas that are responsible for managing the activation and inhibition among the potential candidates (within and between languages), to eventually choose the right language. This makes bilinguals’ use of languages, especially in production, qualitatively different from that of monolinguals. Bilinguals constantly need to check and keep track of the demands of the context they are immersed in and the speakers they are talking to. In addition, once these demands and speakers’ features have been identified, the non-target language has to be suppressed and inhibited to allow for the use of the target language efficiently. Thus, the defenders of the bilingual advantage argue that these additional demands and the extra utilization of inhibitory skills required to efficiently manage the two languages constitute training that would, by transfer, enhance domain general executive function abilities to a higher level in bilinguals. Specifically, it has been argued that inhibition (Bialystok, 2011) and, to a less extent, general response speed enhancement (equated to monitoring, see Costa et al., 2009) would be present in bilinguals. The former would stem from the necessity of inhibiting the non-target language while speaking, while the latter would be a consequence of the constant need to oversee the linguistic demands of the current environment in order to be able to choose the appropriate language for each communicative situation.

The alleged bilingual advantage in inhibitory and monitoring skills has been tested by comparing bilinguals and monolinguals in a variety of classic psychological *tasks* that measure general inhibitory abilities, such as the flanker (Eriksen & Eriksen, 1974), Simon (Simon & Rudell, 1967) and Stroop (Stroop, 1935) tasks. If bilingualism has any enhancing effect on general inhibitory skills, it should be captured by any of those classic tasks that tap into executive functions.

All three tasks include congruent trials (trials where all information presented favours the target response) and incongruent trials (those that information that conflicts with the correct response). In the Stroop task, usually considered the epitome of a task that requires a strong use of inhibitory skills, participants have to name the ink colour of a word presented on the screen (see Fig. 6 for a schematic representation of the task). The word presented on the screen can be the name of the colour the word is printed in (congruent condition, e.g., the word "red" presented in red), a colour word that is different from the colour of the ink (incongruent condition, e.g., the word "red" presented in blue) or a non-colour word (neutral condition, e.g., the word "dog" presented in red). Traditionally, the difference between the congruent and incongruent condition, namely the *Stroop effect*, is taken as an indicator of inhibitory abilities. Crucially, smaller differences have been found between congruent and incongruent trials in bilinguals than in monolinguals (Bialystok et al., 2008).

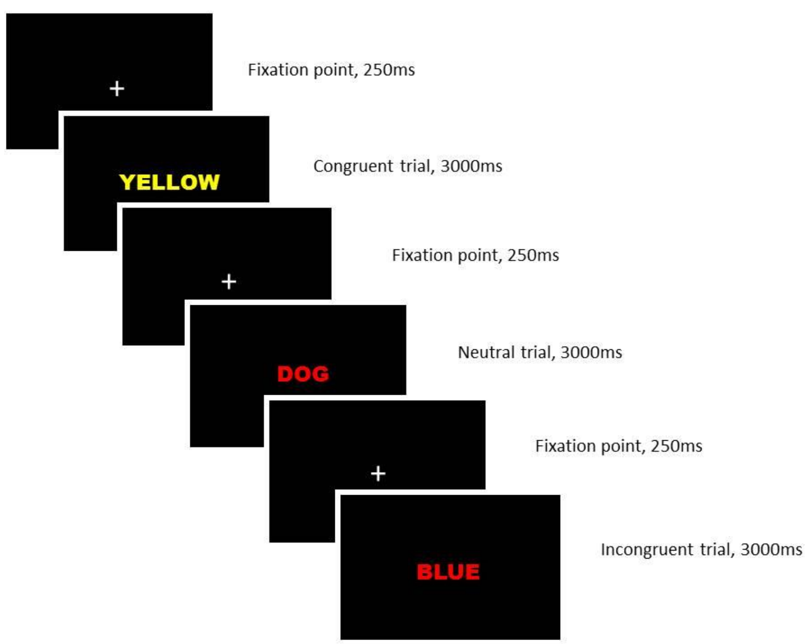

Figure 6. Schematic representation of the Stroop task.

However, considering that this task comprises a strong linguistic component and that it has been shown that bilinguals and monolinguals differ in lexical access (see Ivanova & Costa, 2008; Gollan, Montoya, Fennema-Notestine, & Morris, 2005 and the references mentioned in the second section of the Introduction, "Language organization in bilinguals: the mental lexicon and lexical access"), the conclusions derived from such a task could be questioned. For that reason, researchers generally opted to use tasks that do not engage linguistic information, such as the Simon or flanker tasks.

The Simon task (see Fig. 7) assesses the inhibitory abilities by presenting conflicting spatial information. In this task, participants are presented with geometrical stimuli that contain both an indicator of the response that needs to be given (e.g., press "left" when you see a square and "right" when you see a circle) and position information that is irrelevant for the task, but that can produce incongruent trials (e.g., a circle presented in the left side, indicating that the response "right" needs to be produced) or congruent trials (e.g., a circle presented on the right, the same side as the response that needs to be given). As in the Stroop task, the differences in both reaction times and accuracy between congruent and incongruent trials (i.e., the *conflict effect*) are taken as an indicator of inhibitory skills. Importantly, a smaller *conflict effect* has been found for bilinguals than monolinguals in young adults (Bialystok, 2006), children (Bialystok, Martin, & Viswanathan, 2005) and the elderly (Bialystok, Craik, Klein, & Viswanathan, 2004).

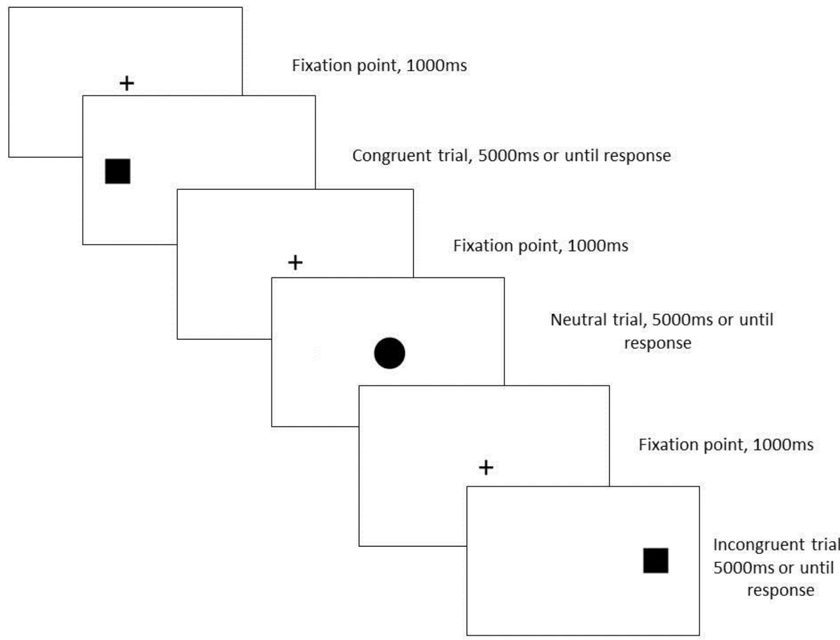

Figure 7. Schematic representation of the Simon task.

Finally, the third most extensively used task to measure inhibitory skills is the flanker task (Fig. 8), in which participants are presented with rows of arrows, and they have to indicate the direction of the central one. The arrangement of the central and the flanker arrows can be such that they create congruent (e.g., ← ← ← ← ←) or incongruent trials (e.g., ← ← → ← ←), and again the difference between these two conditions has been considered as an indicator of inhibitory abilities. Some researchers have found that this *conflict effect* is smaller for bilinguals than for monolinguals (Costa, Hernández, & Sebastián-Gallés, 2008). In sum, the findings that show a reduced *conflict effect* for bilinguals in the Stroop, Simon and flanker tasks have been interpreted as an evidence of a better ability of bilinguals to face incongruent or conflicting situations. Bilinguals are believed to be better at inhibiting the non-target response favoured by the distracting information, thanks to their training in inhibiting the non-target language in daily communicative situations.

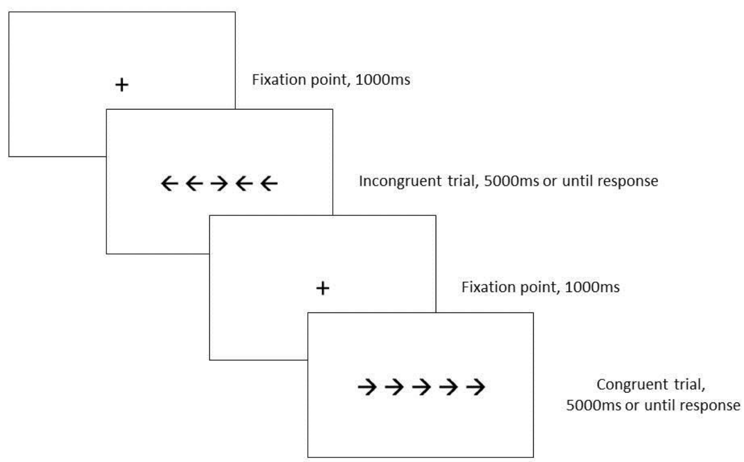

Figure 8. Schematic representation of the flanker task.

Some researchers also found a general response speed enhancement using the same tasks. For example, while testing bilingual and monolingual samples in the flanker task aiming to find a bilingual advantage in conflict resolution (i.e., in inhibitory control), Costa et al. (2008) found that bilinguals were overall faster than monolinguals (see also Martin-Rhee & Bialystok, 2008 and Emmorey et al., 2008). Crucially, this effect was present not only in the incongruent trials, but also in the congruent ones. This result, as they later argued (see Costa et al., 2009), could not be explained by an improvement in the inhibitory capacities – the improvement in the inhibitory skills should affect the participants' responses to the incongruent trials, but not to the congruent trials, where there is nothing to inhibit. As an alternative, the authors claimed that this overall faster performance was due to the better monitoring abilities of the bilinguals, stemming from their expertise in keeping track of the changing linguistic demands in their daily environment and the need to constantly monitor the activation and competition of their language systems. Thus, faster reaction times would supposedly be a reflection of a better ability to monitor a context with changing demands. In line with that, Costa and colleagues (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009) tested bilinguals and monolinguals in flanker task versions that were low-monitoring (which included mostly trials of one kind, either mostly congruent or mostly incongruent) and high-monitoring (more evenly distributed conditions). Although the difference between congruent and incongruent trials (the *conflict effect*) was mostly the same for bilinguals and for monolinguals, they found that the bilinguals were overall faster than monolinguals in the high-monitoring versions of the tasks. This was taken as an indicator of better monitoring abilities in bilinguals, which were only noticeable when the environment was demanding enough (see Bialystok, Craik, Klein, & Viswanathan, 2004 and Martin-Rhee & Bialystok, 2008 for similar results and conclusions). Although this enhancement has been argued to reflect better monitoring abilities, recent views suggest that the equation of overall reaction times on these tasks to monitoring abilities is a very impure approximation that would not account for a clean measure of monitoring (Paap, Johnson, & Sawi, 2014).

These two perspectives, both arguing in favour of a bilingual advantage, are appropriately named by Hilchey & Klein (2011) as "*bilingual inhibitory control advantage*" (or "BICA") and "*bilingual executive processing advantage*" (or "BEPA"), paired to a bilingual advantage in inhibition and in monitoring, respectively. In their review, they show that the bilingual benefits on inhibition are very few, while evidence favouring an overall executive processing advantage is stronger. The advantage in overall reaction times, they argue, is consistent from childhood to old age, and it would stem from an enhanced global conflict monitoring system emerging from the need to monitor linguistic representations competing for selection. As argued before, this general advantage has been equated with monitoring abilities (see Costa et al., 2009; but see Paap, Johnson, & Sawi, 2014 for a critique) and, for the purpose of this thesis they are used interchangeably. However, when Hilchey and colleagues updated their review (Hilchey, Saint-Aubin, & Klein, 2015) with more data, they observed that a general overview of the available research does not support the hypothesis of an advantage in monitoring any more. The following section turns to the criticisms that have been raised to the research practices that found a significant effect reflecting a bilingual advantage.

4. Criticisms to the bilingual advantage hypothesis

a. Hidden demographic factors

In contrast to the accounts that favour 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., intelligence, 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 proposal, which constitutes one of the building blocks of the counter-argument to the so-called bilingual advantage, started with the pioneering work by Morton and Harper (2007). Being aware of the crucial role of 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 demographic factors. Subsequently, other researchers also raised the same objection, 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), intelligence (IQ) and immigrant status. *Higher SES* has been often associated to a better performance in executive functioning tasks (e.g., Mezzacappa, 2004; Noble, Norman, & Farah, 2005) and *IQ* has been positively correlated with EF skills (Adelman et al., 2002; Arffa, 2007). *Immigrant status* and other ethnicity-related factors are known to affect the quality and speed of performance on executive functions as well. Crucially, different countries around the world have immigration policies that seek well prepared newcomers, and consequently the individuals who achieve success in moving from their original country to a new one are usually those with higher IQ (Milne, Poulton, Caspi, & Moffitt, 2001; Wadsworth, Kuh, Richards, Hardy, 2006) which, as explained, is positively correlated with EF. This is of crucial importance when two groups are tested, as is the case of studies on monolinguals and bilinguals. The lack of control might potentially cause differences between groups in those relevant factors, produced by different variations and uneven distributions of the bilingual and monolingual samples.

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. Similarly, 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, & 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.

If the argument of Paap and colleagues (see Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015a; 2015b) has a solid grounding, the bilingual advantage should be reduced or completely eliminated by controlling for several deficiencies that they identify and that may have contributed to the appearance of an 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 participants are carefully matched, any sign for a bilingual advantage, either in inhibition (reduced *conflict effect*) or in monitoring (overall reduced RTs) tends to disappear (see Paap & Greenberg, 2013). Thus, Morton and Harper (2007) tried to replicate the findings obtained by Bialystok, Craik, Klein and Viswanathan (2004) using the Simon task in children, but in contrast 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 the confounding variables were controlled for by matching participant groups and testing large sample sizes (see Paap, Johnson, & Sawi, 2015a). In adequately set experimental conditions, the so-called bilingual advantage systematically vanishes, with monolinguals and bilinguals performing equivalently in samples of children (Gathercole et al., 2014), young adults (Paap & Greenberg, 2013) and the elderly (Kirk, Fiala, Scott-Brown, & Kempe, 2014; de Bruin, Bak & Della Sala, 2015).

b. Developmental factors

While methodological concerns might capture the heterogeneity of the results presented so far in studies comparing monolingual and bilinguals, an additional issue related to the development of certain cognitive skills has recently been raised. Thus, research in this field has mainly focused on studying populations formed of 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*” (i.e., those abilities could be hardly improved more by any other relevant factor) 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. This has been used as a reason of why the bilingual advantage can be hard to capture in young adults and more salient in children and elderly (Bialystok, Craik, & Luk, 2012).

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). In this study, the authors 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 also been questioned recently. For example, 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 carefully matched and large samples, the advantage disappears in child samples (Gathercole et al. 2014). At the other tail of the distribution things are equally unclear. Bilingual seniors' cognitive abilities are already declining due to normal cognitive deterioration caused by age, and thus the problem of the potential "ceiling effect" present in adults should be easily solved and, as a consequence, any potential difference produced by (for example) a linguistic profile difference, easily captured. Some researchers find a bilingual advantage in elderly samples (Bialystok, Craik, Klein, & Viswanathan, 2004; Gold, Kim, Johnson, Kryscio, & Smith, 2013), sometimes even stronger than in the young adults (Bialystok, Craik & Luk, 2008). However, some other studies have recently argued that the evidence indicating better inhibitory skills or general monitoring abilities in elderly bilinguals is not completely reliable and replicable. Namely, Kirk and colleagues (Kirk, Fiala, Scott-Brown, & Kempe, 2014) failed to replicate the results reported by Bialystok et al. (2004) indicating a reduced Simon effect in older bilinguals, and found no indication of overall faster reaction times in older bilinguals either (see also Kousaie & Phillips, 2012a,b; and de Bruin, Bak, & Della Sala, 2015).

c. Sample sizes and replicability

In addition, significant findings for the bilingual advantage happen principally when sample sizes are small (around n<30, see Paap, Johnson, & Sawi, 2015a), confirming the low reliability and replicability of the bilingual advantage effect. Furthermore, significant effects are not always found across the tasks that are assumed to measure the same construct of executive control (Paap & Greenberg, 2013). As Paap and his colleagues argue, for the hypothesis of the bilingual advantage to be coherently demonstrated, the advantage should be present at least in two different tasks that tap into the same cognitive ability and they should correlate, indicating that the same main component underlies the performance of those tasks that set out to measure it (Paap & Greenberg, 2013). Crucially, that is not the case (see, for example, Paap & Greenberg, 2013, Paap & Sawi, 2014). When tasks designed to measure one of the big three components of executive functions (Miyake & Friedman, 2004) are analyzed, surprisingly, different versions of the Stroop task do not correlate with each other (Shilling, Chetwynd, & Rabbitt, 2002). The same is true of different versions of the flanker task (Salthouse, 2010) and of studies comparing performance between different tasks (see Stins, Polderman, Boomsma, & de Geus, 2005; Fan, Flombaum, McCandliss, Thomas, & Posner, 2003; and Kousaie & Phillips, 2012a; for studies showing no or remarkably low correlations between flanker, Simon and Stroop tasks). As a consequence, it is not surprising that studies have shown a bilingual advantage in some tasks but not others (Blumenfeld & Marian, 2014).

Despite the abovementioned inconsistency in the findings, the bilingual advantage hypothesis has been taken as a strong and well-accepted phenomenon. Not surprisingly, the probability of publishing results challenging the potential advantage is usually lesser than the results supporting it. In a recent review and meta-analysis, de Bruin, Treccani, and Della Sala (2014) analyzed more than 100 conference abstracts exploring the bilingual advantage, and they reported that half of them supported it and the other half partially or completely challenged it. Importantly, most of the studies used similar tasks, and the ones showing an advantage and the ones challenging it did not differ in statistical power or sample size. Crucially, the authors observed that while 68% of the conference abstracts that showed a bilingual advantage were published, only 29% of the studies showing a bilingual disadvantage or no differences between bilinguals and monolinguals were published. The authors took this difference as an indicator of a publication bias.

The issues exposed above highlight the importance of exploring the bilingual advantage meticulously, give credit to both significant and null results, and interpret what has been accepted as common wisdom under the light of scientific data.

VI. The purpose of this thesis

If the use of two languages implies a more extensive and intense use of domain-general cognitive abilities such as executive functions, which have been shown to be improved by training, it seems reasonable to assume that said functions would be enhanced as a consequence of the intense language control that a bilingual life requires. Nevertheless, the evidence in this regard is far from clear. Many articles and researchers argue against this advantage that is, according to them, a product of extraneous non-linguistic factors that were not controlled for. The contradictory evidence, therefore, draws an unclear picture regarding the alleged bilingual advantage and additional studies are required to better understand whether bilingualism has any positive impact on executive functions.

The present thesis is an attempt to shed light on this issue. I will explore the reliability of the bilingual advantage phenomenon, in both inhibitory and monitoring abilities, on the three main age groups in which it has been reported using large samples of children, young adults and elderly. To do so, the main concerns raised by the opponents to the bilingual advantage (Paap & Greenberg, 2013) will be accounted for, by testing large samples in the main tasks that have been used when significant results were reported, and by controlling for any external variable that could affect the outcomes and their interpretation. Also, the reliability of the performance in different tests (see Paap & Greenberg, 2013, Paap & Sawi, 2014; for arguments against cross-task consistency) will be analyzed by computing correlations comparing conflict indexes across tasks that the same individuals undergo.

This exploration will be performed by testing bilingual participants from the Basque Country, a bilingual community in the north of Spain and the south of France. As reported by the Basque Language Academy (Euskaltzaindia), Basque became co-official (together with Spanish) in 1980. Since then, Basque has been present in most spheres of the society, from government bodies to schooling, culture or media, and this presence is increasing over the decades. This situation permits bilingual people living in this community to use both languages freely while they are exposed to them on a daily basis. Thus, as opposed to bilingual individuals from other communities that might use their languages with context restrictions (e.g., one at work, one at home), Basque bilinguals can use both languages without restraints.

If the bilingual advantage is real, it should be captured in (at least one of) the tasks that are going to be used in the present thesis. If the reason for its elusive nature is the ceiling effect that is present in the young adults, it should be easily detected in children and elderly. However, if the previously found significant results are a consequence of uncontrolled external factors that are unevenly distributed across bilinguals and monolinguals, both groups should behave similarly in every task run in every group of the present thesis.

VII. *Structure of the thesis*

This thesis is divided in 5 main chapters: the General Introduction, three experimental chapters (one exploring the effects in seniors, one in children and one in young adults), and the General Discussion. While the first and the last ones constitute the theoretical framework of the work presented here, the three experimental chapters describe the experiments conducted in order to address the issues raised in the introduction, as well as the conclusions derived for each of them.

Chapter 1 frames the concept of the "bilingual advantage" and the need of shedding light on its legitimacy. It starts by contextualizing the concept of bilingualism by firstly exploring language organization and language use in monolingual speakers, after which those same aspects are analyzed in detail in the case of a bilingual. The mechanisms for language control are explored, emphasizing the relation between the two languages in bilinguals and the ways in which they can affect each other. Once the language system is described, different consequences of bilingualism are discussed. First, the linguistic perspective is adopted to see the possible disadvantages and advantages that bilingualism can provide in relation to language knowledge and management. Second, cognitive consequences are described, including the recent evidence on cognitive disadvantages and advantages that have been directly associated to the knowledge of a second language. Among the latter, I delve into the "bilingual advantage", understood as bilinguals' advantage in EF. I explain its origins and different interpretations, and then present the empirical data that support it. Similarly, addressing the ongoing debate that surrounds the existence of such an advantage, I describe the recent critiques that this hypothesis received. Considering the unclear picture arising from this unsolved debate, I end the introduction by proposing a developmental exploration of the existence of a bilingual advantage in large and well matched samples of monolinguals and balanced native bilinguals, arguing that even though young adults are the most explored population in the field, seniors and children should be more likely to show any potential effect of bilingualism on EF.

Chapter 2 is focused on the effects of bilingualism in seniors. Its introduction reviews what we know about the existence of a bilingual advantage in elderly samples, as well as the purported protective nature of bilingualism for dementia symptoms. Being aware of the methodological critiques that previous research has received, I present data of carefully examined senior samples in four experiments. In the first two, the performance of equivalent samples of bilinguals and monolinguals in the verbal and numerical versions of the Stroop task is analyzed. Similarly, in the third and fourth experiments, the same tasks are used to analyze the performance of a heterogeneous group of bilinguals who vary in their second language knowledge. Individuals' performance across task is also analyzed to obtain indexes of between-task correlations. The discussion of the chapter summarizes the findings, interprets them and points to the need of exploring the same hypothesis in different age groups.

Chapter 3 explores the reliability of the bilingual advantage in children. The introduction of this chapter explores the recently reported data on its presence and absence in different contexts, as well as the present knowledge about the normal development of the EF in monolingual children. After that, I present data from very large children samples of different ages from bilingual and monolingual schools. The children had the same characteristics, and they took part in the same tasks used for the seniors in Chapter 2, and therefore the same analysis are conducted to explore any between group difference and cross-task coherence. Furthermore, the ANT task was added to have a stronger measure of any possible difference. The discussion presents the implications of the results, both in terms of the development of EF in bilingual children and in any potential difference that it could have been found between them and their monolingual peers.

Chapter 4 closes the developmental circle by focusing on young adults. It first introduces the existing evidence on the bilingual advantage in adults, by far the most studied population in this topic. Furthermore, it presents the data that comes from testing young adults in all the tasks used in previous chapters, as well as an additional task that has been extensively used in the literature that allows for a more extensive and detailed cross-task exploration. The performance of the tested adults is summarized and interpreted in the discussion, first focusing on the between group results and then comparing the results obtained in adults to those of children and seniors in Chapters 2 and 3.

For the sake of clarity and comprehensibility, the results from the different tasks, as well as any correlation between the tasks used, will be concisely summarized in a table at the end of each chapter.

In Chapter 5, I first review the results obtained in the experimental chapters, and I compare them to the previous research presented in the introduction, to try to find explanations for the differences between the results presented here and previously reported findings. Different explanations and hypotheses are proposed that could account for the data presented here and that could open new ways of exploring this phenomenon in future research.

Chapter 2: the bilingual advantage hypothesis in the elderly

I. Overview and theoretical introduction

The bilingual advantage hypothesis is usually considered to correspond to a general enhancement of cognitive control, mainly due to the bilinguals' constant inhibition of the non-target language (Green, 1998). The inhibition occurs because both languages that the bilinguals speak are always active (Kroll, Dussias, Bogulski, & Kroff, 2012) and competing for selection. For that reason, bilinguals have to use their executive functions (especially inhibition and, to a less extent, monitoring) more than monolinguals. Bilingual advantage defenders argue that this constitutes a very efficient training that eventually leads to an enhancement of domain general executive functions. Consequently, it sounds plausible to assume that the more somebody trains a given function, the better at it he eventually becomes. Following that stream of reasoning, the more time somebody has been a bilingual, the greater her enhancement of executive functions should be with respect to a monolingual peer. If bilingualism has any enhancing effect on general cognitive skills, they could be expected to be more salient after a whole life of exposure to this training – i.e., in the elderly, rather than young bilinguals. Potentially, this training could also slow down or protect against the normal decline associated to aging that the cognitive abilities suffer (see next section).

1. Previous evidence on the bilingual advantage in seniors

Unfortunately, there are very few studies providing seemingly consistent evidence favouring a behavioral advantage in tasks measuring different forms of cognitive control and executive functions in elderly and lifelong 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). However, 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, supporters of the bilingual advantage have also argued in favour 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 & Sebastián-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. 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* does indeed exist, but only in people with low educational level, where higher degrees of bilingualism were associated with a delay in the diagnosis. As a potential solution for the conundrum brought by the hazy role of bilingualism in patients with dementia, a handful of studies have opted for a longitudinal approach, testing cohorts from the 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). Even though testing this assumption is not the purpose of the present thesis, it is worth noting that the cognitive benefits in the elderly bilinguals are under debate in areas other than the executive functions.

2. Aim of the chapter

These inconsistent pieces of evidence make it difficult to determine whether lifelong bilingualism truly has any beneficial effects on executive functions in elderly bilinguals or not. It seems to be the accepted view that the use of two languages yields boosting or enhancement of domain-general cognitive abilities. If this is so, that enhancement should be easily captured in elderly lifelong bilinguals, who have been actively exposed to two languages for decades, consequently training their inhibitory (according to the "BICA" perspective, see Hilchey & Klein, 2011) and monitoring (the "BEPA" perspective, Hilchey & Klein, 2011) skills for much longer than younger bilinguals. 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 on cognitive reserve in the elderly. The first set of experiments presented in the current chapter (Experiments 1 and 2) is aimed at investigating any relevant differences in cognitive control between elderly bilinguals and monolinguals in either inhibition or monitoring, when they are compared on a verbal and non-verbal Stroop task.

Considering that it has been suggested that bilingualism should be treated as a continuous rather than dichotomous factor (e.g., Singh & Mishra, 2013), the second set of experiments (Experiments 3 and 4) explores the impact of the "degree of bilingualism" - here measured as non-immigrant bilinguals' general proficiency in their second language - on executive control functions. A wide range of bilingual seniors with varying degrees of second language (L2) proficiency were tested in the verbal Stroop (Experiment 3) and the numerical Stroop tasks (Experiment 4), to see if the classic markers of inhibition (as well as monitoring) are modulated by participants' L2 knowledge and mastery.

In both of the experimental tasks, correlations between the indexes obtained in both Stroop tasks are used to see whether the same inhibitory processes are being applied to both tasks or if, on the contrary, the cross-task reliability is low, as suggested by some critics (Paap & Greenberg, 2013, Paap & Sawi, 2014).

II. Experiment 1: Effect of lifelong bilingualism in the verbal Stroop task

In this first experiment, the hypothesis of the existence of bilingual advantage in inhibitory skills was tested by comparing elderly bilinguals' and monolinguals' performance in a verbal Stroop paradigm (Stroop, 1935). The goal was to explore whether bilinguals' and monolinguals' inhibitory skills and/or monitoring abilities differ in any way.

The verbal Stroop task is one most popular and most widely studied tasks used to measure inhibitory control in general population as well as in bilinguals (MacLeod, 1991). If, according to the bilingual advantage hypothesis, bilinguals outperform monolinguals in tasks that require inhibition of irrelevant information, it could be expected that bilinguals perform much better than monolinguals on this task. In fact, this has been 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) . They found a larger *Stroop effect* (the difference between the congruent, i.e. the word "red" presented in red, and incongruent trials, i.e. the word "red" presented in green) for older monolinguals as compared to their bilingual peers. Furthermore, they wanted to explore in detail how the responses to congruent and incongruent trials contributed to the general *Stroop effect* difference. Hence, they analyzed the congruency effects (i.e., the difference between the RTs to the congruent trials, i.e. "red" presented in red, and the RTs to neutral symbol trials, i.e., a string of dollar signs printed in red) and the *incongruity effects* (i.e., the RTs in incongruent trials, meaning the time taken to name the colour of a word such as "red" printed in blue, compared to the time needed to name neutral symbols). They found that the congruency effect was larger and that the *incongruity effect* was smaller in the bilingual sample than in monolingual sample (see Hernández et al., 2010, for a similar pattern). 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 non-linguistic skills is primarily evident in advanced stages of life (e.g., Bialystok, Craik, & Freedman, 2007; see Hilchey & Klein, 2011, for review).

It has been recently argued (Paap, Johnson, & Sawi, 2015a) that the studies reporting bilingual advantages are confounded by the uncontrolled extraneous factors. For instance, the majority of the bilingual sample in the study by Bialystok et al. (2008), as well as many other studies that showed a bilingual advantage (e.g., Bialystok, 1986; Bialystok & Senman, 2004; Bialystok & Shapero, 2005; Bialystok, 1999; Bialystok, & Martin, 2004) consisted of immigrant individuals, a status which can affect performance of executive functions. Due to the immigration policies of several countries, people who successfully manage to move from one country to the other are usually those with higher IQ (Milne, Poulton, Caspi, & Moffitt, 2001; Kuhn, Everett, & Silvey, 2011; Wadsworth, Kuh, Richards, & Hardy, 2006), and higher IQ is usually associated with a better performance in executive functions (Adelman et al., 2002; Arffa, 2007). Therefore, some studies argue that, once the confounding external factors such as immigration are controlled for, the differences between bilinguals and monolinguals in executive functions cease to exist (Paap & Greenberg, 2013).

The results of the present experiment (verbal Stroop) are going to be explored by the classical hypothesis testing (by comparisons employing ANOVAs) but, given that this approach does not allow for accepting the null hypothesis, the critical differences of interest are also tested following the Bayesian Null Hypothesis Testing approach (see Kruschke, 2013; and Rouder et al. 2009, among others) by obtaining the Bayes Factor (BF). The Bayes Factor has a comparative nature, and it is an index obtained when the likelihood of the data is considered under the assumption that the null hypothesis is true and compared to the likelihood of the data under the assumption that the alternative hypothesis is true. The Bayes factor must be understood as a ratio that reflects the likelihood of one hypothesis over the other (Jarosz & Wiley, 2014). For example, a BF_{01} (indicating that the null hypothesis, i.e. 0, is tested against the alternative, 1) of 6 indicates that the data is 6 times better explained under the assumption of the null hypothesis being real than under the assumption of the alternative hypothesis. Hence, the general guideline to interpret the results of this statistic analysis is that it should be taken as the likelihood to accept the first model over the second one. That is why there is not a standard value (as it occurs with the p-values) according to which one could accept one of the hypotheses over the other. Instead, it is taken as a ratio. However, some authors have established that, the first model can be accepted if the resulting BF is above 3 and the second one if the resulting BF is below 0.3, (see Krushchke, 2011, Fig. 3 in page 6; but see Raftery, 1995 and Jeffreys, 1961, for slightly different boundaries).

Both groups of participants were carefully matched on 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 in elderly to low 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, the following groups of well-controlled participants should still show significant differences either in the magnitude of the *Stroop effect* (i.e., bilinguals should show a reduced *Stroop effect* 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). If, on the other hand, the previous significant effects were a consequence of external factors, once these are controlled for, both participant groups should behave similarly.

1. Method

a. 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 research center).

Half of the participants (out of which 14 were female; mean age= 68.75, SD=4.42) were Spanish monolinguals, and the other half (out of which 14 were female, mean age of 69.38, SD=4.58) were Basque-Spanish bilinguals who use both languages every day 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 self-rated 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 did not differ in any demographic factor (all *p*s>.5, see Table 1), including the age at which they ceased formal schooling, the IQ percentile based on the scores 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), participants from all the educational strata were recruited, equally distributed among both language groups: 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 did not differ significantly in general, speaking and comprehension abilities in Spanish (all *p*s >.6, see Table 1 for detailed information), which was the language in which they were spoken to and tested in during the whole process.

Table 1.- Characteristics of the samples of monolingual and bilingual seniors tested in Experiments 1 and 2. Mean values for each group are displayed with standard deviations between brackets. *P* values 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	(.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	(.90)	8.79	(1.06)	***0.88***
Basque general	--	--	8.04	(.95)	--
Basque speaking	--	--	8.13	(1.08)	--
Basque comprehension	--	--	8.29	(1.37)	--

b. Materials

For the *verbal Stroop* task (which was a variation of the classic Stroop task; Stroop, 1935), the names of the colours red, blue and yellow ("rojo", "azul" and "amarillo" in Spanish) and three pairwise-matched non-colour 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 colour name printed in the same colour that it indicates, e.g., the word "rojo" in red), incongruent (a colour name printed in a different colour, e.g., the word "rojo" in blue) and neutral word (words that were not colour names printed in the three colours, e.g., the word "ropa" in red) conditions. Also, neutral symbol (a string of dollar signs printed in the three colours, e.g., "$$$$$" in red) condition was added to create a condition unaffected by language. Each condition consisted of 24 trials, and each colour was presented the same amount of times in each condition, associated the same amount of times to each written word. Each participant was presented with a total of 96 experimental trials, and the trial presentation order was randomized across participants.

c. Procedure

All the participants were tested in the BCBL facilities in Donostia-San Sebastián. The experiment was run using DMDX (Forster & Forster, 2003). 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 colour 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. The participants first saw a fixation mark for 250ms and then the target word appeared on the screen for 3000ms. All the strings were presented in uppercase Courier New font on a black background, and the colours were set in the RGB-scale values as follows: blue=0,0,255; red=255,0,0; yellow=255,255,0.

2. Results

a. Latencies

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 4 x 2 ANOVA was run including the factors Condition (congruent, incongruent, neutral words and neutral symbols) and Group (monolinguals and bilinguals). In this general

ANOVA, a significant effect of Condition was found [$F(3, 136) = 66.88$, $p<.01$] but neither the main effect of Group nor the interaction resulted significant (all *F*s<1). All the descriptive values of conditions in reaction times and accuracy across groups are detailed in Table 2, as well as the effects obtained from those conditions.

Table 2.- Mean latencies for correct responses and error rates in all conditions for the verbal Stroop task for the monolingual and bilingual groups of seniors in Experiment 1. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets.

		Reaction times				**Error rates**			
		Monolinguals		**Bilinguals**		**Monolinguals**		**Bilinguals**	
Conditions	**Congruent**	772	(217)	787	(167)	0.52	(1.41)	0.87	(1.73)
	Incongruent	1017	(202)	1001	(176)	2.26	(4.07)	2.95	(5.83)
	Neutral Word	901	(185)	892	(131)	0.69	(2.01)	0.69	(2.01)
	Neutral Symbol	791	(144)	780	(105)	0.35	(1.18)	0.17	(.85)
	Total	871	(167)	865	(130)	0.95	(1.53)	1.17	(2.09)
Effects	**Stroop**	246	(167)	213	(128)	1.74	(4.06)	2.08	(5.63)
	Congruency	129	(164)	104	(128)	0.17	(2.60)	-0.17	(2.29)
	Incongruity	-117	(104)	-109	(125)	-1.56	(2.96)	-2.26	(4.91)

To explore all the possible venues in which differences could emerge, 2 x 2 ANOVAs were also computed to analyze the potential differences between the groups in the *classic Stroop effect* (i.e., congruent vs. incongruent), *incongruity effect* (neutral word vs. incongruent conditions) and *congruency effect* (neutral word vs. congruent conditions)[1]. These differences were further tested 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), by testing the H_0 of no differences against

[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].

the H_1 according to which 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 Fs<1, see Fig. 9), 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* effect, with significant main effect [$F(1, 46)= 46.42$, $p<.01$), but no effect of Group or an interaction (all Fs<1), and no group differences when the *incongruity effect* was 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: $BF_{01}>5.04$).

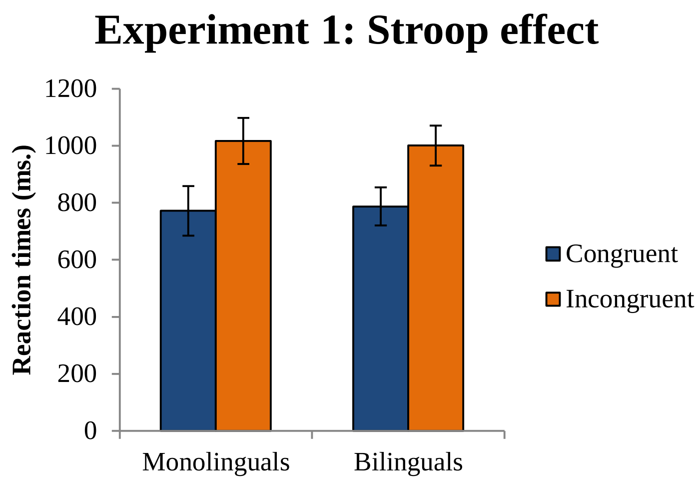

Figure 9. Reaction times for congruent and incongruent conditions in Experiment 1. Error bars represent confidence intervals of 95%

b. Accuracy

The error rate analysis also showed a quite similar pattern, first with a general 4 x 2 ANOVA and then comparing each index separately by using 2 x 2 ANOVAs and Bayesian Null Hypothesis Testing between groups (see Table 2). The general 4 x 2 ANOVA showed a significant effect of Condition [$F(3, 138)= 8.15$, $p<.01$], but no main effect of Group and no interaction (Fs<1). In the analysis of the *Stroop effect*, a main Condition effect is observed [$F(1, 46)=7.27$, $p<.02$] but neither the main effect of Group nor the interaction was significant (Fs<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, no effect of Condition, Group or an interaction was found (all Fs<1), and Bayesian analysis comparing the index across groups 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 ($Fs<1$), and a Bayes factor analysis favoured the null hypothesis over the alternative (Bayesian t-test of the index between groups: $BF_{01}>3.92$).

Hence, every analysis conducted on the data from Experiment 1 unambiguously indicated that bilingual and monolingual participants performed similarly in the verbal Stroop task. The general ANOVAs and the individual index comparisons in both latencies and accuracy failed at showing any significant difference in performance and, crucially, the Bayes factor favored the null hypothesis of no differences in every comparison.

III. Experiment 2: Effects of lifelong bilingualism in the numerical Stroop task

Even though the results from the classic Stroop task are very robust and consistent, it is important to be aware of an inherent problem of the Stroop task when dealing with language-based test groups. Namely, the classical version of the task necessarily involves linguistic stimuli, which makes it admittedly difficult to isolate differences in pure inhibitory skills from differences driven by basic linguistic performance variations (linked to participants' linguistic skills and proficiency), and this is of crucial importance due to the fact that bilinguals and monolinguals regularly display differences in the time needed to complete lexical access (e.g., Bialystok et al., 2008; Ivanova & Costa, 2008, see "The negative linguistic impact" in the 4th section of the Introduction).

For this reason, researchers investigating the relationship between executive control and multilingualism have recently adopted different approaches to the systematic investigation of the differential executive function-related effects in monolinguals and bilinguals using non-linguistic tasks. In order to be able to explore the real impact of bilingualism in inhibitory skills, another similar Stroop-like non-linguistic task that mainly taps into inhibition and that doesn't require spoken responses was included: The numerical Stroop task. This task, also called 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).

This allows for generalization of the results obtained from the first experiment or, if the new results challenge the old ones, for reinterpretation of the data and the assumptions originally made. If the null effects obtained in the previous task were a consequence of lexical access or spoken word production differences between bilinguals and monolinguals, those disadvantages should play a less important role in this task and thus differences in EF abilities between bilinguals and monolinguals should be observed. On the contrary, if the results indicating similar performance of both linguistic groups are strong and replicable and not dependent on lexical access variations, the same pattern should be obtained when the same participants are tested in the verbal and non-verbal Stroop task.

1. Method

a. Participants

The same participants that took part in the Experiment 1 were tested in the Experiment 2 (see Table 1).

b. Materials

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 of pairs of 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 number larger in value was also bigger in size, e.g., small digit 1and big digit 6), 16 incongruent trials (the number larger in value was smaller in size, e.g., big digit 1and small digit 6) and 16 neutral trials (two equal numbers, different in size, e.g., big digit 1-small digit 1). In all the conditions, “left” and “right” responses were equally distributed.

c. Procedure

All the technical equipment and software used in this experiment were identical to that reported for the verbal Stroop task in Experiment 1. On 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, with each digit of each pair on one side of the screen.

2. Results

The same rationale as in the previous experiment was followed in the analysis of the *numerical Stroop* task results.

a. 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 Fs<1]. See Table 3 for detailed information for each condition.

Table 3.- Mean latencies for correct responses and error rates in all conditions for the numerical Stroop task for the monolingual and bilingual groups of seniors in Experiment 2. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets.

		Reaction times				Error rates			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	Congruent	615	(133)	608	(111)	0	(0.00)	0.26	(1.28)
	Incongruent	693	(153)	684	(133)	2.08	(5.43)	2.34	(4.81)
	Neutral	621	(125)	631	(111)	0	(0.00)	0.52	(1.76)
	Total	643	(131)	641	(114)	0.69	(1.81)	1.04	(1.94)
Effects	Stroop	77	(86)	76	(60)	2.08	(5.43)	2.08	(4.76)
	Congruency	5	(37)	24	(38)	0	(0.00)	0.26	(5.04)
	Incongruity	-72	(90)	-52	(73)	-2.08	(5.43)	-1.82	(1.28)

Following the procedure applied in Experiment 1, in the current experiment the critical *classic Stroop effect* (i.e., comparing the congruent and the incongruent conditions) was explored together with the more detailed *incongruity* (responses to neutral vs. responses to incongruent conditions) and *congruency effects* (neutral vs. congruent conditions)[2] using both ANOVAs and Bayesian Null Hypothesis Testing to compare the indices between groups. *Stroop* (see Fig. 10) and *incongruity* effects showed the same pattern, with the strong Condition effects (all *ps*<.01) and no other main effects or interactions (all *F*s<1), while Bayesian analysis also showed that the Null Hypothesis of no differences was more likely than the alternative 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 Condition was significant [F(1, 46)=7.25, p<.02], but no effect of Group (F<1) and no interaction [F(1, 46)=2.91, p>.1] were found, although the results from the Bayesian Null Hypothesis testing did not allow to reach any conclusion (Bayesian t-test of the index between groups: BF_{01}>0.58).

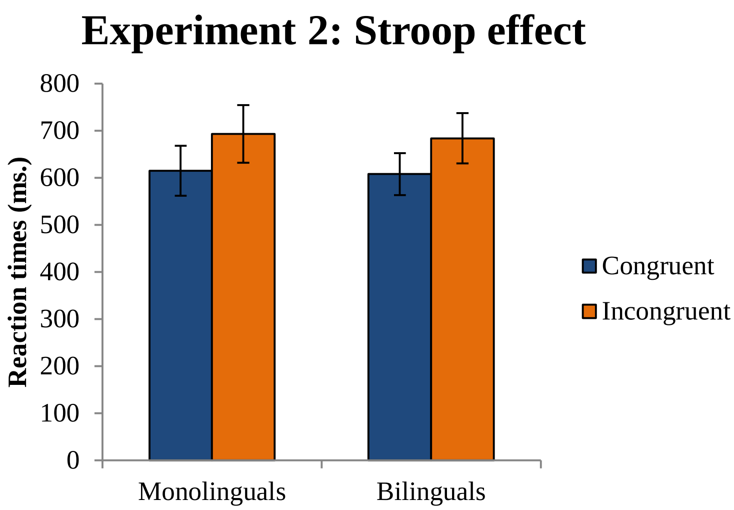

Figure 10. Reaction times for congruent and incongruent conditions in Experiment 2. Error bars represent confidence intervals of 95%

b. Accuracy

In a similar vein, the general 3 x 2 ANOVA on the error rates (see Table 3) 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, the *Stroop effect* was significant [F(1, 46)= 8.00, p<.01] but no main effect of Group nor an interaction was found (all

[2] As in Experiment 1, 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 *ps*>.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 *ps*>.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 *ps*>.14].

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], nor an interaction between them was found [F(1, 46)= 1.87, p>.18]. Moreover, 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).

As in Experiment 1, the analyses performed to explore the results from the numerical Stroop task indicated a very similar performance of both bilinguals and monolinguals. The general ANOVAs and the individual index comparisons in both latencies and accuracy did not show any significant different between groups and, importantly, the Bayes factor favored the null hypothesis over the alternative in most of the cases. Crucially, it never favored the alternative one.

IV. Interim conclusion: Experiments 1 and 2

Results of the two tasks in the first two experiments 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 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 assumption 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. Even though most of the participants tested were not highly educated, this was not the case for all of them. Considering that the differential effects could be stronger in low-educated samples (see Gollan et al., 2011), an additional set of analyses was run including only the participants with the lowest education levels, which provided with parallel results (see footnotes 1 and 2).

Furthermore, the cross-task coherence between the two Stroop tasks for elderly bilinguals and monolinguals was analyzed by computing correlations between the indices obtained (*Stroop, congruency* and *incongruity*). The analyses indicated that the *Stroop effect* (r= -.11), *incongruity effect* (r= .26) and *congruency effect* (r= .10) showed very mild cross-task correlations (for similar results, see Paap & Greenberg, 2013, Paap & Sawi, 2014).

The results I presented go against the previously reported pieces of evidence indicating that senior bilinguals outperform their monolingual peers in executive control tasks (Bialystok, Craik & Luk, 2008; Bialystok, Craik, Klein, & Viswanathan, 2004; Gold, Kim, Johnson, Kryscio, & Smith, 2013), but go in line with the recent results reported by other researchers arguing that these significant effects were not reliable and replicable (Kirk, Fiala, Scott-Brown, & Kempe, 2014; Kousaie & Phillips, 2012; de Bruin, Bak & Della Sala, 2015).

However, there is a different question that might arise when discussing the bilingual advantage, and that could have not been properly captured in the previous task: instead of a clear cut dichotomous distinction between monolinguals and bilinguals, it has been argued that L2 proficiency could modulate the effect (Singh & Mishra, 2013; Tse & Altarriba, 2014; Goral, Campanelli & Spiro, 2015, see below for further clarifications). If the potential benefit in executive functions is modulated by proficiency, this difference might not be captured by simple comparisons between monolinguals and bilinguals. To investigate this potential modulation, another experiment was conducted. It was based on the same tasks, but a different sample was under the scope: a set of bilingual participants who differ in L2 proficiency, ranking from a very limited knowledge to perfectly fluent and balanced bilinguals.

V. *Experiment 3: effect of the L2 proficiency on the Verbal Stroop task in lifelong bilinguals.*

It has been suggested that instead of focusing on comparisons between bilinguals and monolinguals, bilingualism should be better treated as a continuous rather than dichotomous factor (Valdés, 2001) and thus the effects of bilingualism should differ with a level of bilingualism. 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. Similarly, Tse and Altarriba (2014) found modulations based on L2 proficiency in the Simon task. However, when Goral, Campanelli and Spiro (2015) tested dominant and balanced bilinguals ranging from adults (50 years) to seniors (84 years) in different tasks (including Simon task), they found that balanced bilinguals showed an age-related inhibition decline (the older they were, the bigger the *conflict effect* was) while dominant bilinguals (with lower L2 proficiency) showed smaller or no decline at all. On the other hand, some other researchers have also considered this hypothesis and found no positive or negative modulations in inhibitory control, monitoring or switching based on participants' L2 proficiency (see, for example, Paap, Johnson, & Sawi, 2014). In the second sets of experiments that are presented in the current chapter, the possibility that the proficiency in the second language modulates the bilingual advantage (if any advantage is revealed as significant) is explored by testing a group of bilinguals that vary highly in their L2 proficiency from almost monolingual to perfectly fluent bilingual.

1. Method

a. Participants

70 Basque-Spanish bilingual seniors (including the 24 bilinguals tested in Experiments 1 and 2) were recruited from the same city in the Basque Country (Donostia-San Sebastián) and were non-immigrants (45 females; mean age of 69.36, SD=4.40; age range = 61-81). 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 4 for detailed information about participants' profiles). None of them reported history of chronic neuropsychological disorders 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 Experiments 1 and 2, 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 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 4.- Characteristics of the samples of bilingual seniors tested in Experiments 3 and 4. Mean values are displayed with standard deviations between brackets.

	Bilinguals	
Chronological age (years)	69.36	(4.40)
Education (age)	17.71	(4.71)
MMSE (raw score)	29	(1.30)
General IQ (percentile)	70	(29.65)
Spanish general	8.72	(1.08)
Spanish speaking	8.65	(1.07)
Spanish comprehension	8.99	(.99)
Basque general	6.49	(2.40)
Basque speaking	6.7	(2.62)
Basque comprehension	7.23	(2.13)

b. Materials

Materials used for this experiment were the same as the ones used in the Experiment 1.

c. Procedure

The procedure followed was the same as the one in the Experiment 1.

2. Results

The same analysis as the one used in Experiment 1 was performed here, but instead of a between-subject factor separating bilinguals from monolinguals in two groups, the Basque General Proficiency was considered as a covariate. This rating scale varied from 1 (very poor level of Basque) to 10 (very fluent) among the 70 bilingual speakers.

a. Latencies

Firstly, 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 5 for descriptive values).

Table 5.- Mean latencies for correct responses and error rates in all conditions for the verbal Stroop task for the groups of seniors in Experiment 3. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets.

		Verbal Stroop task			
		Reaction Times		**Error Rates**	
Conditions	**Congruent**	789	(145)	0.42	(1.26)
	Incongruent	1011	(174)	2.14	(4.42)
	Neutral Word	891	(140)	0.36	(1.37)
	Neutral Symbol	765	(104)	0.18	(.85)
	Total	864	(128)	0.77	(1.46)
Effects	**Stroop**	223	(107)	1.726	(4.29)
	Incongruity	-120	(98)	-1.79	(4.11)
	Congruency	103	(108)	-0.06	(1.66)

After this general analysis of the 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 a covariate[3]. A significant effect of Condition [$F(3, 204)= 18.29$, $p<.01$] was found, but the effect of Basque General Proficiency was not significant [$F(1, 68)=2.11$, $p>.15$], and it did not interact with Condition ($F<1$).

[3] Considering previous findings that show a relation between the intelligence scores and the different executive functioning tasks such as the Stroop task (Adelman et al., 2002; Arffa, 2007), a four-way ANCOVA including both Basque General Proficiency and IQ percentile values (obtained from the K-BIT) as covariates was also run. 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 [$F(3, 201)= 7.29$, $p<.01$]. 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 Condition and IQ[$F(1, 67)=15.25$, $p<.01$], indicating that the *Stroop effect* was smaller for higher IQ values ($r= -0.44$, $p<.01$, $n= 70$). A similar pattern is observed 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 Fs<1). Basque General Proficiency factor was not significant (all Fs<1, all ps>.33) for any of the indices, nor did it interact with Condition.

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) 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 [$F(1, 68)=1.20$, $p>.28$].

b. Accuracy

In the error rate analysis, the general four-way ANCOVA showed that none of the effects or interactions were significant [all Fs<2 and all ps>.17, see Table 5].

c. Additional analysis: Educational level

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 that the so-called *bilingual advantage* has been claimed to depend on this factor (cf. Gollan et al., 2011), an additional analysis was run 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 Fs<1]. When each index was explored independently in the corresponding set of two-way ANCOVAs, the *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 [$Fs(1,67)<2.6$, ps>.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 same four-way ANCOVA run on the accuracy data showed that none of the effects or interactions were significant (all *F*s<1.8 and *p*s>.2). Altogether, the marginal main effects of Education that emerged out of 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.

d. Additional analysis: correlation between L2 proficiency and indices

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 Fig. 11). The error rate analysis showed the same pattern, and neither the *Stroop* ($r=.04$, $p>.74$), nor the *congruency* ($r=-.05$, $p>.66$) nor the *incongruity* ($r=.06$, $p>.64$) indices were correlated with the general proficiency that participants had in Basque (see Footnote 3).

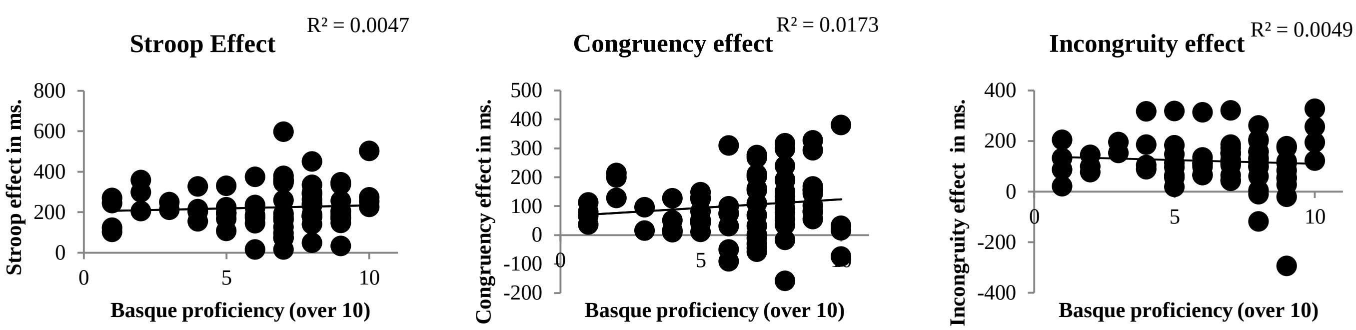

Figure 11. Correlations between the indices obtained in Experiment 3 by each participant and their proficiency in Basque. Proficiency is indicated by participants' self-reported values (from 0 to 10). Indices reflect RT differences between congruent and incongruent (*Stroop effect*), congruent and neutral (congruency effect) and incongruent and neutral (*incongruity effect*). R^2 value is indicated for each index.

As in Experiment 1, the classic indices from the verbal Stroop task were significant. However, none of these indices were modulated by the proficiency in participants' L2, Basque, and these results did not change when the education level was considered as a covariate. Importantly, the indices were not correlated with Basque proficiency either, indicating that the level of bilingualism does not significantly affect the performance in a task like the verbal Stroop, which heavily relies on general executive functioning abilities.

VI. *Experiment 4: effect of the L2 proficiency of lifelong bilinguals in the numerical Stroop task.*

Given that many studies have shown that bilinguals might suffer a disadvantage in production of spoken responses (Ivanova & Costa, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005), the numerical Stroop task (Besner & Coltheart, 1979; Kaufmann et al., 2005; Santens & Verguts, 2011) was used with the same set of participants that conducted the Experiment 3. This was in line with the rationale followed in the Experiments 1 and 2.

1. *Method*

a. *Participants*

The same participants that were tested in Experiment 3 took part in this experiment (see Table 4).

b. *Materials*

Materials used for this experiment were the same as the ones used in the Experiment 2.

c. *Procedure*

The procedure followed was the same as the one used in the Experiment 2.

2. *Results*

a. *Latencies*

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 6 for the descriptive results). 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 *Stroop* [$F(1, 68)=12.54$, $p>.01$] 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 Fs<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 6.- Mean latencies for correct responses and error rates in all conditions for the numerical Stroop task for the groups of seniors in Experiment 4. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets.

		Numerical Stroop task			
		Reaction Times		Error Rates	
Conditions	Congruent	598	(103)	0.45	(1.62)
	Incongruent	671	(117)	1.79	(4.00)
	Neutral	621	(100)	0.27	(1.28)
	Total	630	(103)	0.83	(1.60)
Effects	Stroop	73	(56)	1.34	(4.37)
	Incongruity	-50	(61)	-1.52	(3.90)
	Congruency	24	(34)	-0.18	(1.83)

b. Accuracy

The general three-way ANCOVA run in error rates showed that neither Condition [$F(2, 136)=.17$, $p>.84$], neither Basque General Proficiency, nor their interaction was significant (Fs<1, see Table 6).

[4] As in Experiment 3, a general ANCOVA that included both Basque General Proficiency and IQ percentile values as covariates was conducted. 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 between Condition and IQ[$F(1, 67)=5.78$, $p<.02$], showing that the *Stroop effect* decreased as IQ values increased (r= -0.28, $p<.02$, n= 70). In the *congruency effect* a significant IQ effect was found [$F(1, 67)= 10.89$, $p<.01$], with no other significant main effects or interactions (all Fs<2, all ps>.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 Fs<1 and ps>.53).

c. Additional analysis: Educational level

As in the Experiment 3, 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 [$F<1$], as well as no interaction between Condition and Basque General Proficiency [$F<1$], or 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].

d. Additional analysis: Correlation between L2 proficiency and indices

Additionally, a 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, 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 Fig. 12). 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*: $r=.09$, $p>.44$; see Footnote 4).

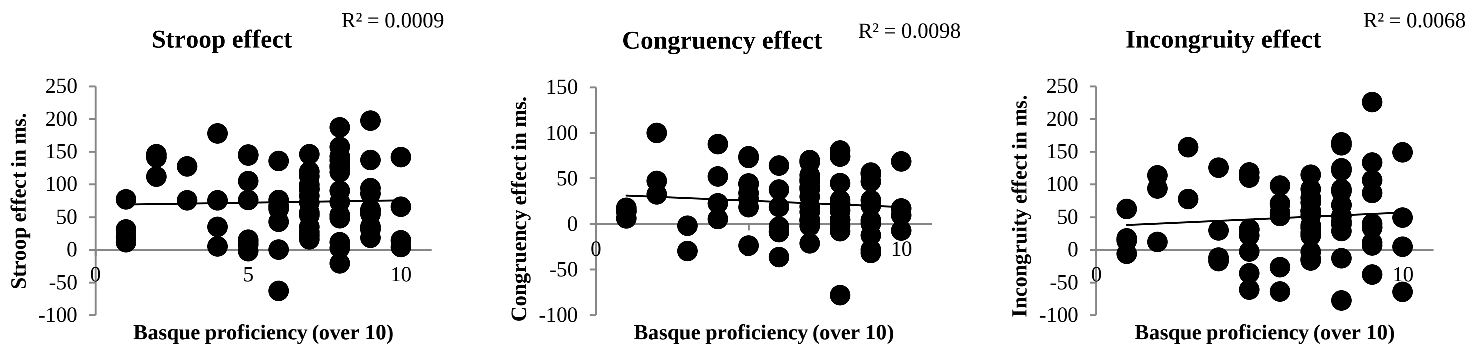

Figure 12. Correlations between the indices obtained in Experiment 4 by each participant and their proficiency in Basque. Proficiency is indicated by participants' self-reported values (from 0 to 10). Indices reflect RT differences between congruent and incongruent (*Stroop effect*), congruent and neutral (*congruency effect*) and incongruent and neutral *(incongruity effect*). R^2 value is indicated for each index.

As in Experiment 3, none of the classic indices from the numerical Stroop task were significantly modulated by the proficiency in the participants' L2, Basque. These findings did not change when the variability coming from the educational level was also accounted for. Remarkably, Basque proficiency and the classic indices did not show any significant correlation, indicating once again that the level of bilingualism does not significantly change the performance on general executive functioning tasks like the numerical Stroop.

VII. *Interim conclusions: Experiments 3 and 4*

In Experiments 3 and 4, significant and strong *Stroop effects* were obtained in the latency analysis of both verbal and numerical Stroop task, mainly due to the *incongruity effects*. However, when the impact of Basque General Proficiency in the different indices was analyzed, and even when factors such as IQ or Educational Level were controlled for, the ANCOVAs showed no significant effect of participants' knowledge of a second language or a modulation of the main indices based on the level of this knowledge, as measured by their Basque General Proficiency.

Preceding research on this issue has failed to provide a consistent picture. Thus, to the best of my knowledge, the current study is the first one that aimed at checking for a possible modulation of bilingual seniors' inhibitory capacities by the degree of their L2 mastery, controlling for other factors. To investigate this potential modulation, I used the same tasks as in Experiments 1 and 2, but with a sample of bilingual participants who differ in

L2 proficiency, ranging from a very limited knowledge to perfectly fluent and balanced bilinguals. 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 the 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 subsequent correlations demonstrated that, regardless of their L2 proficiency, participants showed comparable inhibitory skills (as measured by the *Stroop effects*), thus extending the earlier evidence showing lack of a significant relationship between L2 proficiency and the inhibitory control measures (Paap, Johnson, & Sawi, 2015a). When the ANCOVAs were run including the IQ percentiles obtained from the K-BIT as a covariate, previous findings were replicated, suggesting that the *Stroop indices* are reduced for higher IQ values (see Adelman et al., 2002; Arffa, 2007). Still, no main effect of Basque General Proficiency was observed, 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 no significant difference in the executive functions as a function of their L2 knowledge was observed in the elderly lifelong bilinguals.

Furthermore, and in line with what was shown in Experiments 1 and 2, the cross-task coherence between the two Stroop tasks was tested by running a correlation analysis on the indices obtained by the bilingual seniors on both tasks. The results indicated that the *Stroop effect* (r= .02) and the *congruency effects* (r= .34) showed a very mild cross-task correlations, while *incongruity effect* was significantly – but negatively – correlated across tasks (r= -.52).

VIII. General discussion: bilingual and monolingual seniors

In the Experiments 1-4, the effects derived from lifelong bilingualism on domain-general cognitive abilities related to inhibitory control and monitoring skills were analyzed in the samples of elderly bilinguals and monolinguals.

In Experiments 1 and 2, Spanish speaking elderly monolinguals and Spanish-Basque elderly bilinguals who have been immersed in a bilingual society and who have been using both languages for the vast majority of their lives on a daily basis were compared. However, it has been suggested that the L2 proficiency modulates the effect (Singh & Mishra, 2013; Tse & Altarriba, 2014; Goral, Campanelli & Spiro, 2015), an issue that was explored in Experiments 3 and 4. In these two experiments the experimental sample consisted of a large group of bilinguals that differed in their L2 proficiency, ranging from the bilinguals with low knowledge of the second language to perfectly fluent and balanced bilinguals. Under the assumption that any potential impact of bilingualism on cognitive functioning should be modulated by the degree of knowledge of the second language, any effects obtained in verbal and numerical Stroop task should show a significant correlation with seniors' L2 proficiency.

All of the tested participants were non-immigrants coming from the same city and did not differ in any of the demographic factors, nor in 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 the participants scored above 26 in Experiment 1 (Median = 29.5), and above 24 in Experiment 2 (Median = 29). As they have been exposed to bilingualism their whole life and as they are not at their peak of cognitive abilities, these bilinguals were tentatively selected as a good test case to explore any enhancing effect that bilingualism may have on the inhibitory control and monitoring skills. If bilingualism provides individuals with any protecting, boosting or enhancing features regarding their cognitive abilities, any difference should be easier to be captured in a sample that is ongoing a normal declining process, as is the case in the one tested here.

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 *Stroop effects*), and importantly, these effects were not modulated within a bilingual group by varying knowledge of their L2. This provides evidence that that the level of knowledge and use of bilingualism does not have a direct impact on general cognitive abilities. 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.

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. 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. when comparing bilinguals and monolinguals. In the present study investigating Basque-Spanish bilinguals and Spanish monolinguals, for example, one could argue that Basque speakers represent a cultural minority with a different social and historical background compared to those of Spanish monolinguals. However, it is important to mention in that regard that the participants come from the same city and that, not being immigrants, they share cultural and historical background to a great extent. Therefore, I think that the samples are as comparable as a between-subjects design allows for.

The results found here add to the growing body of evidence that has been gaining strength in the last several years and that suggests that the bilingual advantage in executive functioning (and explicitly in inhibitory and monitoring abilities) is actually non-existent (Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015; Paap, & Sawi, 2014). However, I acknowledge that the present results should be considered with caution. Although the 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 (such as 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 benefit of bilingualism when the relevant 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 Woumans et al., 2015). However, as correctly stated by Paap, Johnson and Sawi (2015b), the most compelling pieces of evidence regarding this may come from longitudinal studies following cohorts of individuals. Indeed, most of the longitudinal studies yielded non-significant differences, or even monolingual-favouring 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 study presented evidence in favour of a bilingual advantage at this level of analysis (Wilson et al., 2015).

In summary (see Table 7), the present data allows to conclude that the bilingual seniors (of the characteristics tested in the current experiments) do not benefit from their knowledge of a second language, when they are tested in domain-general executive control tasks. However, one might argue that this difference might be present in the samples with other characteristics. Critically, the bilingual and monolingual seniors might display certain individual differences, such as cognitive impairments and other relevant factors related to aging, that might obscure the potential benefits derived from bilingualism. Therefore, to test whether the null effects obtained here were a consequence of age-related issues, the same hypothesis was explored in the complete opposite tail of the distribution: comparing bilingual and monolingual children.

Table 7.- Summary of the results obtained in the present chapter. A cross indicates non-significant effects of language in the measures obtained. Correlations are reported only if significant.

	Tasks used	Inhibition	Monitoring	Between-task correlation
Elderly	*Verbal Stroop*	✕	✕	Incongruity in Exp. 3-4, r= -.52
	Numerical Stroop	✕	✕	Rest, n.s.

Chapter 3: The bilingual advantage hypothesis in children

I. Overview and theoretical introduction

In the second chapter I showed how bilingualism did not have any beneficial (nor detrimental) impact on elderly participants' executive functions. However, testing elderly participants involves complications that can be difficult to account for. For example, some participants may have mild cognitive deteriorations that remain undetected during the analysis. Also, the individuals might differ in their linguistic and other life-long experiences, which can be hard to quantify. For these reasons, in this Chapter I will test the same hypothesis that was tested in the previous Chapter, but in children.

1. Previous evidence on the bilingual advantage in children

As was commented in the Introduction, the bilingual advantage should be equally apparent in children as it is in the elderly. It has been suggested (Bialystok, Craik, & Luk, 2012) that the lack of a bilingual advantage in young adults does not prove its absence, but simply implies that it might be captured more reliably in the extreme tails of the demographic distribution, such as in children. Generally, when bilingual advantage is explored, as occurs in most psycholinguistic research, young adults are the usual participant group profile that researchers have access to. Importantly, this participant group happens to be at the peak of their domain-general cognitive abilities (20-40 years of age, see Hartshorne & Germine, 2015). Therefore, it would not be surprising that this peak in their abilities results in a ceiling effect and camouflages any potential bilingual advantage effect that might otherwise be detected. This idea is not new and some studies have provided supporting evidence: Bialystok, Martin and Viswanathan (2005) tested 5 year-old children, young adults (20 years of age), middle-aged adults (30-59 years of age) and older adults (60-80 years)from both bilingual and monolingual populations in the classic Simon task, and found that bilinguals outperformed monolinguals in the groups of children and middle-aged and older adult groups, but that any trace of bilingual advantage was virtually absent in the young adult participants.

The inability to find evidence of a bilingual advantage is not limited to the young adult age group: the previous literature suggests that the effect also seems elusive in research with children and the number of conflicting studies of this type has increased in recent years. One the one hand, some studies show the advantage for bilingual children as compared to their monolingual peers (e.g., Martin-Rhee & Bialystok, 2008; Yang, Yang, & Lust, 2011). For example, general findings from experiments investigating task switching speak in favour of benefits of bilingualism in young participants (4-5 years old). For instance, the results from classical card sorting tasks, where participants are asked to sort the cards by one feature (e.g., by their colour) in some trials, and to ignore that feature and to sort them by another feature (e.g., by their shape) in other trials, showed that bilingual children perform better than monolingual ones in the second sorting process (i.e., after shifting from classifying based on one set of features to the other set). The cost of switching from one set of rules to the other (reflected in longer time needed or more errors committed), i.e. the switch cost, was smaller for bilinguals, suggesting that bilingualism improves task switching (e.g., Bialystok & Martin, 2004; Bialystok, 2001; Craik & Bialystok, 2006) and putatively provides support for a bilingual advantage. On the other hand, however, recent findings suggest that, using carefully matched and large samples of children, this advantage disappears (as happened with the elderly sample). As an example, Gathercole et al. (2014) tried to account for the problems that Paap and Greenberg (2013) brought up, noting that many of the studies showing a bilingual advantage could possibly be showing a Type I error, due to inadequately matched or very small groups, uncontrolled external factors or task-dependency effects. In order to do so, Gathercole and colleagues tested a large number of Welsh children and adults from the same sociocultural background in different tasks (n=650 in a card sorting task, n=557 in the Simon task and n=354 in a grammaticality judgment task). The different groups tested included English monolinguals and bilinguals with different degrees of use of Welsh and English (i.e., bilinguals who only spoke Welsh at home, bilinguals who used both Welsh and English at home, and bilinguals from English-speaking homes). Importantly, Gathercole et al. found no evidence for a bilingual advantage. No differences were found in the switch cost or overall performance in the card sorting task. Similarly, negligible differences were found in the Simon task. The grammaticality judgment task also failed to reveal any systematic bilingual advantage. Thus, the picture regarding bilingual advantage in children is not straightforward.

Interestingly, it should be mentioned that even researchers showing differences between bilingual and monolingual adults in EF have sometimes admitted that the evidence in favour of a bilingual advantage in children is certainly limited (Bialystok, Barac, Blaye, & Poulin-Dubois, 2010; Bialystok, Craik, & Luk, 2012; see also Hilchey & Klein, 2011, for a review).

When exploring the potential differences between two groups of children in a construct such as EF, it is important to bear in mind that these functions in this age range are under the process of development, and that process needs to be understood in order to explore possible between-group effects. The development of EF has been explored in different ways (see Anderson, 2002; and Jurado & Rosselli, 2007; for a review) and it is assumed that, developmentally speaking, it is one of the slowest cognitive abilities. The reason behind this is most probably that these functions are associated with the prefrontal cortex (PFC), which has a protracted maturation (Diamond, 2002). A commonly accepted perspective is that EF components reach adult-like level by the end of childhood or early adolescence (Diamond, 2002; Welsh, 2002). Concretely, updating or working memory capacities have been shown to develop through childhood into adolescence (Brocki & Bohlim, 2004; Beveridge, Jarrold, & Pettit, 2002; Gathercole, Pickering, Ambridge, & Wearing, 2004); task shifting shows comparable-to-adult behaviors in children around the age of 12 (Cepeda, Kramer, & Gonzalez de Sather, 2001; Kray, Eber, & Lindenberger, 2004) and similarly, inhibition abilities increase through childhood until reaching adult-like levels somewhere between the age of 12 and early adolescence (Klenberg, Korkman, & Lahti Nuuttila, 2001; Durston et al., 2002; Van den Wildenberg, & Van der Molen, 1995; Williams, Ponesse, Schachar, Logan, & Tannock, 1999). To explore the age-related changes of the three main components of the EF, Huizinaga and colleagues (Huizinga, Dolan, & van der Molen, 2006) tested children of 7, 11 and 15 years of age and adults of 21 years of age in tasks tapping into these three main components, and found that, in line with previous findings, different components mature at different moments: inhibitory abilities developed rapidly by the age of 11, shifting continued developing into adolescence and working memory continued until young-adulthood.

One of the goals of studying EF functions in bilingual and monolingual children is to observe whether their developmental trend, explained in the previous paragraph, varies depending on the linguistic profile. This is in order to see whether the so-called bilingual advantage impacts significantly on the development of the EF functions – especially on inhibitory or monitoring abilities. One could tentatively argue that the fact that adult and elderly monolinguals and bilinguals perform similarly does not necessarily mean that there are no EF differences during their development, as the path to achieve the same mastery in EF might differ for bilinguals and monolinguals.

As mentioned above, however, it has been suggested that some factors other than the mere linguistic profile of the participants (monolinguals vs. bilinguals) may play a very important role in the emergence of the bilingual advantage in different tasks. For instance, Morton and Harper (2007) tested a group of bilingual and monolingual children in a Simon task and they found no differences in their performance as a function of the number of languages they spoke. Instead, they found a significant correlation between their socio-economic status (SES) and their performance in the task, arguing that the SES, not bilingualism, was the crucial factor in producing the effect. Hence, there is a number of external factors that seem to have a direct impact on the appearance (and the magnitude) of the bilingual advantage, and the true nature of the outperformance of the bilinguals on executive control tasks remains dubious, casting doubts on some of the claims that have led the field in the last decade.

2. Aim of the chapter

As was explained in the introduction and in the opening section of this Chapter, the likelihood of finding differential effects for bilingual advantage is greater in children than in adult samples, given that children are far from having fully developed inhibitory skills and consequently they are expected to be more sensitive to the difference between congruent and incongruent trials. Furthermore, studies show that the inhibitory skills reach full development at some point between 11-12 years and early adolescence (Huizinga, Dolan, & van der Molen, 2006; Klenberg, Korkman, & Lahti Nuuttila, 2001; Durston et al., 2002; Van den Wildenberg, & Van der Molen, 1995; Williams, Ponesse, Schachar, Logan, & Tannock, 1999). Following these findings, the participants tested in the present chapter range between 7 to 14 years of age, the critical developmental period in which the EF, especially inhibition (as it is the main component of interest), develop prolifically. Nevertheless, as full development of the cognitive abilities is not yet reached in this age range (Hartshorne & Germine, 2015), no ceiling effects that could mask any potential effect of bilingualism are expected. The experiments performed with bilingual seniors (Chapter 2) were based in the same rationale: the potential effects stemming from bilingualism would not be camouflaged by ceiling effects because seniors' cognitive abilities would already be in decline due to normal aging. However, one could arguably claim that the possible effect of bilingualism was not captured in seniors because precisely this cognitive decline equalized seniors' EF abilities. Children, then, seem to be the perfect sample to test the bilingual advantage hypothesis and its impact on the development of the EF: in this population, cognitive skills are still not fully developed, but they are not yet in the process of decline. Thus, if bilingual advantage truly exists, it could be captured in the present chapter in two different ways: it could result in enhanced inhibitory skills, as a reduced *conflict effect* (i.e., the *Stroop effect*), in line with the "BICA" hypothesis (Hilchey and Klein, 2011), or in enhanced monitoring skills, as a reduction of the overall RTs, in line with the "BEPA" hypothesis (Hilchey & Klein, 2011).

Recent proposals have drawn attention to the fact that various external factors (such as SES) might influence the results in the bilingual advantage tasks (Paap, Johnson & Sawi, 2015). Therefore, all the potential affecting factors will be controlled for in the experiments in this Chapter. In this way, it should be possible to detect, as finely as possible, any possible impact that bilingualism might have in the emergence and development of the executive functions. Importantly, considering the idiosyncrasy of the bilingual educational system in the Basque Country, a relatively high degree of control of children's use of the two languages can be applied. Simply by checking their academic syllabus and the language in which each subject is taught, daily exposure to both languages can be ensured. If the previously reported bilingual advantage was due to those external factors, no differences in either monitoring or in inhibitory abilities should be observed between the bilingual and monolingual samples.

Given the need for methodical investigation of the bilingual advantage in children, I explore this issue in three experiments: in the first two experiments (Experiment 5 and 6), bilingual and monolingual children went through both verbal and non-verbal versions of the Stroop tasks, using the same paradigms as the tasks described in the previous chapter. To that end, large samples of more than 250 bilinguals and 250 monolinguals of different ages were recruited at different elementary and high schools. In the third experiment (Experiment 7), a group of 360 children (180 bilinguals, 180 monolinguals) were compared by means of a child-friendly version of the ANT (see Rueda et al., 2004), to investigate whether there is a bilingual advantage in children in any of the attention networks, and whether the development of these networks is similar or different for bilingual and monolingual children.

II. Experiment 5: the verbal Stroop task in children.

As explained in the Introduction and in the previous Chapter, the Stroop task (Stroop, 1935) is one of the most popular and most widely-studied tasks that has been used to measure inhibitory control. Even though it has been well studied and established in young adults and elderly, when the Stroop task has been tested in children with the purpose of testing the bilingual advantage, an inconclusive pattern is observed. When exploring the bilingual advantage in preescholers, some authors found discrepancies in the results depending on the task applied. Thus, bilingual preschoolers have been shown to display an advantage over monolinguals on conflict resolution tasks such as Simon (Martin-Rhee, & Bialystok, 2008), Dimensional Change Card Sort (Bialystok & Martin, 2004) and the ANT (Yang, Yang, & Lust, 2011). However, this difference is not found when age-appropriate Stroop tasks are used (Martin-Rhee & Byalistok, 2008; Siegal, Iozzi, & Surian, 2009). Children-adapted Stroop tasks tend to be univalent, meaning that only one kind of information is presented to kids in each trial. For example, in the case of the Day/Night task (Gerstadt, Hong, & Diamond, 1994), kids are asked to say “day” when they are presented by the picture of a moon, and “night” when they are presented with a picture of a sun. These adapted Stroop tasks, although they tap into the inhibition of the dominant responses and the ability to face incongruent situation, do not contain any distracting information that requires suppression (see also Archibald & Kerns, 1999 and Livesey, Keen, Rouse, & White, 2006; for other variants of Stroop adaptations to children), and that is why, arguably, bilinguals show no differences with regard to their monolingual peers (see Martin-Rhee & Bialystok, 2008). Proving this hypothesis, Esposito and colleagues (Esposito, Baker-Ward, & Mueller, 2013) tested preschool children in the Day/Night task, but also in a bivalent Stroop-like shape task, which was based on stimuli containing both relevant and distracting information that needed to be inhibited. Critically, they found a bilingual advantage only in the latter, where distracting information had to be inhibited (see also Carlson & Meltzoff, 2008).

The goal of this experiment is to test whether the results obtained in adults with the Stroop task (Bialystok et al., 2008) but that were not replicated in the elderly (cf. Chapter 2) can be found in children. Furthermore, being aware of the limitations of running Stroop-like tasks in preschool children (Esposito et al., 2013), the classic verbal Stroop task was run with children that already acquired literacy, and thus the same task that has been shown to capture bilingual advantages in adults is going to be used in the present sample, to see whether children vary in their monitoring or inhibitory abilities as a function of their linguistic profile. Nevertheless, if the lack of consistency across the studies testing monolingual and bilingual children reflects that the significant effects were a product of small sample sizes and inadequately matched samples (see Paap & Greenberg, 2013; Paap, Johson, & Sawi, 2015), null differences should be expected in well controlled large populations. Crucially, both groups of participants were carefully matched in the potentially confounding factors that were discussed through the introduction (cf. Paap & Greenberg, 2013), which includes matching them in general intelligence, immigration and ethnicity, among others, so that the only relevant difference between both groups was their linguistic profile.

As in the previous Chapter, the results will be firstly tested with the classical approach (by comparisons employing ANOVAs) but the critical differences of interest will be also tested following the Bayesian Null Hypothesis Testing approach (see Rouder et al. 2009; Krushchke, 2011, among others).

1. Method

a. Participants

Two groups of participants were recruited from 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 three, four, five and six of different elementary schools and from grades one and two of different high schools (42 participants from each grade). The mean age of each grade can be found in Table 8. These Spanish-speaking monolinguals had no fluent knowledge of any other language and were recruited from different schools, from those Spanish provinces where Spanish is the only official language. None of them corresponded to any immigrant minority. Furthermore, they were exclusively exposed to Spanish at home, as indicated by 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 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 and present in everyday life. All the bilingual children attended bilingual schools where the two languages were used as vehicular languages in the educational practice, i.e. where bilingual linguistic educational model was used. 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). The bilingual participants were carefully selected to ensure that all of them were born in the Basque Country and that none of them corresponded to any specific minority group. Bilingual and monolingual participants were carefully matched in different measures and cognitive skills (see Table 8 for detailed information).

Table 8.- Characteristics of the bilingual and monolingual children tested in Experiments 5 and 6 divided by grade. Mean values are displayed with standard deviations between brackets.

		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)	
Primary School	**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)
3rd grade	**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	**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)
4th grade	**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	**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)
5th grade	**Bilinguals**	10.1	(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	**Monolinguals**	11	(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)
6th grade	**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	**Monolinguals**	12.1	(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)
1st grade	**Bilinguals**	12.1	(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	**Monolinguals**	12.9	(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)
2nd grade	**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***	
	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)
Total	**Bilinguals**	10.5	(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***	

Except for 1 bilingual participant, the parents of all the participants in the study reported Spanish as their first language (L1). Note 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 non-academic 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).

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, non-verbal 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 vs. bilingual) levels, but also at the individual grade level, as shown in Table 8. None of the t-tests resulted significant (all *p*s>.35 at the group level and all *p*s>.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, as well as the teachers). Second, none of the participants 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, all of the 504 participants scored above the 20th centile on the verbal, non-verbal and composed IQ tests. Hence, given the careful matching of the participants, I consider that little doubt could be cast on similarity between test groups regarding all the factors except for the linguistic profile.

b. Materials

Eight Spanish words were used in the classic verbal Stroop task: the names of the colours, 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 colour 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 was created by presenting each of the colour names printed in the colour that matched the lexical entry (e.g., the word "verde" printed in green ink). The incongruent condition was created by presenting each colour name printed in a colour that did not match the colour represented by the lexical entry (e.g., the word "verde" printed in red ink). The neutral word condition was created by presenting the non-colour words in the ink colour that corresponded to their pairwise-matched counterparts from the colour name set. In all the conditions, each word was presented six times, paired equally to each colour, resulting in a total of 24 trials per condition. Finally, a control symbol condition was also included 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 ("%%%%%") were presented in the four possible ink colours in a total of 24 trials. Hence, each participant was presented with a total of 96 experimental trials. The trial presentation order was randomized across participants.

c. Procedure

The students were tested in their schools, using the same technological equipment for 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 the data was gathered during school hours, in specific dependencies that the schools kindly provided for this purpose. Participants were first presented with a recording of the instructions via headphones. They were instructed to name the colour 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 of the instructions. Following this procedure, the potential impact of experimenter-driven differences in the recording sessions was minimized. 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. The participants first saw a fixation mark that was briefly displayed in the center of the screen for 250ms 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 2500ms. All the strings were presented in uppercase Courier New font on a black background. The precise RGB-scale values for each of the colours 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 minutes.

2. Results

Individual verbal responses were collected and resulting data were preprocessed and corrected for incorrect voice key triggering with CheckVocal (Protopapas, 2007). Incorrect responses (less than 2% of the data) and reaction times below or above 2.5 standard deviations from the mean in each condition for each participant (less than 2.5% of the data) were excluded from the latency analysis. The mean latencies for correct responses and error rates are presented in Table 9. After the general 4 x 2 ANOVA (Condition x Language), different ANOVAs were conducted in order to explore the classical *Stroop effect* (incongruent vs. congruent trials), *the incongruity effect* (incongruent vs. neutral word trials), and 3) *the congruency* (congruent vs. neutral word trials)[5], as well as their interaction with Language. These differences were further tested 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 for *Stroop, congruency* and *incongruity effect* of the two groups). This way, the H_0 (no differences between the indices for monolingual and bilingual group) was tested against the H_1 (smaller indices for bilinguals than monolinguals), examining the so-called bilingual advantage with Bayesian t-tests.

[5] A parallel set of analysis was performed 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 9), it was 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, it is worth stressing 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: Fs<1.17, ps>.28; error rates: Fs<2.20, ps>.14].

Table 9.- Mean latencies for correct responses and error rates in all conditions for the verbal Stroop task for the groups of children in Experiment 5. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets. Effects are shown underneath the conditions.

		Reaction times				Error rates			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	Congruent	771	(137)	784	(136)	0.89	(2.61)	0.48	(2.26)
	Incongruent	963	(176)	977	(181)	4.6	(6.41)	5.04	(7.43)
	Neutral Word	870	(163)	892	(167)	1.54	(3.84)	1.82	(4.24)
	Neutral Symbol	769	(129)	781	(127)	0.83	(2.26)	0.69	(2.34)
	Total	841	(140)	855	(140)	1.96	(2.81)	2.01	(3.18)
Effects	Stroop	191	(107)	193	(105)	3.7	(6.27)	4.56	(7.02)
	Congruency	98	(93)	108	(99)	0.64	(3.8)	1.34	(3.78)
	Incongruity	-93	(86)	-85	(86)	-3.06	(5.82)	-3.22	(6.47)

a. Latencies

First, a series of ANOVAs including the factors Condition (4 levels: congruent, incongruent, neutral words, neutral symbols) and Language (2 levels: monolinguals, bilinguals) were conducted to see whether the different levels of Condition behaved differently and whether Language modulated it. A significant effect of Condition was found [$F(3,1506)=1058.17$, $p<.01$], but the effect of Language was negligible [$F(1,502)=1.47$, $p>.23$]. The interaction between these two factors did not approach significance [$F<1$]. Next, each of the individual effects[6] were explored (see Table 9).

[6] To check for any modulations of the effect and its development across the six different grades and two language profiles, the ANOVAs were repeated including Grade as a factor. Analysis of the RT on the *Stroop effect* showed a main effect of Grade [$F(5,492)=23.64$, $p<.01$], but no other effect or interaction [Fs<1 and *ps*>.71], demonstrating a larger effect in younger than in older participants, but similar effect for both bilinguals and monolinguals. Analysis of the error data showed a parallel pattern, with a main effect of Grade [$F(5,492)=17.67$, $p<.01$], and no effect of Language [$F(1,492)=2.447$, $p>.11$], nor interaction [$F<1$, $p>.59$]. *Incongruity effect* was not modulated as a function of Grade [$F<1$, $p>.88$] or Language [$F(1,492)=1.01$, $p>.31$], and interaction between these factors was not found [$F(5,492)=1.20$, $p>.30$]. Accuracy analysis showed a main effect of Grade [$F(5,492)=7.59$, $p<.01$], but no significant Language effect or interaction [all Fs<1]. 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. *Congruency* analysis on RT showed an effect of Grade [$F(5,492)=31.15$, $p<.01$], 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. Error rate analysis of congruency showed an effect of Grade [$F(5,492)=7.53$, $p<.01$], 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 [$F<1$, $p>.65$].

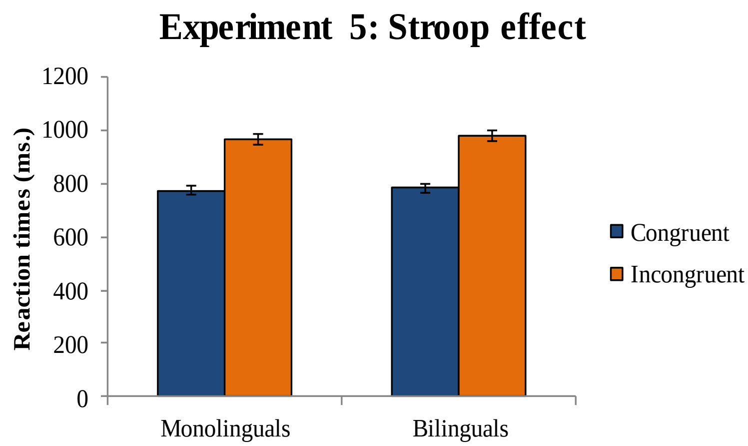

Figure 13. Reaction times for congruent and incongruent conditions in Experiment 5. Error bars represent confidence intervals of 95%

To explore the *Stroop effect,* a series of ANOVAs including the factors Condition (incongruent vs. congruent) and Language (bilingual vs. monolingual) was run. In the reaction time analysis, a significant effect of Condition was found [$F(1,502)=1663.65$, $p<.01$], showing that the incongruent trials were responded to slower than the congruent ones. Language [$F(1,502)=1.07$, $p>.30$] and its interaction with Condition [Fs<1 and ps>.83] were not significant (see Fig. 13). When the hypothesis of bilinguals and monolinguals behaving dissimilarly was tested using the Bayesian approach, the results unambiguously supported the null hypothesis of no differences ($BF_{01}= 9.89$).

A different series of analyses was performed in order to explore the *incongruity effect* (incongruent vs. neutral word). The neutral word condition was used as a baseline condition, instead of the neutral symbol condition, given that different processing biases related to the processing differences between linguistic and non-linguistic materials wanted to be avoided. ANOVAs on the reaction times revealed a main effect of Condition [$F(1,502)=539.91$, $p<.01$], but no effect of Language [$F(1,502)=1.58$, $p>.20$], but no interaction between Language and Condition [$F(1,502)=1.02$, $p>.31$]. Bayesian factor analysis clearly favoured the null hypothesis as the best fit to the data ($BF_{01}= 6.16$).

In order to explore the facilitation caused by the *congruency effect*, a two-way ANOVA was conducted, 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 [$F(1,502)=582,86$, $p<.01$]. However, the main effect of Language was not significant [$F(1,502)=1.90$, $p>.16$], nor was the interaction between Condition and Language [$F(1,502)=1.29$, $p>.25$]. There was a significant *congruency effect* (shorter reaction times for congruent stimuli than for neutral stimuli), but this difference was similar for bilinguals and monolinguals. The *congruency effect* was similar for the two Language groups, as indicated by the conservative Bayesian t-test analysis that favoured the null hypothesis (BF_{01}=5.39).

b. Accuracy

To explore the *Stroop effect*, the ANOVA including Condition (incongruent vs. congruent) and Language (bilingual vs. monolingual) was run on error rates. It revealed a main effect of Condition [$F(1,502)=194.45$, $p<.01$], showing that participants made more errors in the incongruent than in the congruent trials (see Table 9). However, the Language effect was not significant [Fs<1]. The Language by Condition interaction was also not significant [$F(1,502)=2.10$, $p>.14$]. Again, the Bayesian approach favoured the null hypothesis when bilinguals and monolinguals are compared (BF_{01}= 3.64).

When the *incongruity effect* (incongruent vs. neutral words) was explored, ANOVAs on the error rates revealed a significant main effect of Condition [$F(1,502)=131,31$ $p<.01$], but no effect of Language or interaction [Fs<1 and ps>.39]. The *incongruity effect* were virtually identical for the two Language groups, as reported by the Bayesian factor analysis (*BF01*= 9.66).

The *congruency effect* was analyzed on the error rates by running ANOVA including Condition (congruent vs. neutral word) and Language as factors. The ANOVA showed a main effect of Condition [$F(1,502)=34.84$, $p<.01$], showing that participants made more errors in the neutral condition than in the congruent condition. The main effect of Language was not significant [Fs<1 and ps>.78]. However, the interaction between Language and Condition was significant [$F(1,502)=4.22$, $p=.04$]. To better understand the origin of this interaction, each language group was analyzed separately. The effect of Condition was larger for bilinguals than for monolinguals (bilinguals: [$F(1,251)=31.58$, $p<.01$], monolinguals: [$F(1,251)=7.25$, $p<.01$]. Expectedly, the Bayesian analysis did not unambiguously support any of the hypothesis (BF_{01}= 1.30), and therefore I should withhold any final conclusions.

Thus, the results observed in the Stroop task closely resemble what was obtained with seniors in Chapter 2. Significant *Stroop effects* were obtained, mainly produced by *incongruity effects*, and these indices were equivalent across groups. Both the classic ANOVA analysis and the Bayesian null hypothesis testing revealed that there were no signs of bilingual advantage whatsoever, even when the variability coming from the age group was included in the analysis. However, even though the *congruency effect* was comparable in monolinguals' and bilinguals' latencies, it did differ in accuracy, showing a larger effect for bilinguals. Although this larger effect was significant, it is difficult to conclude that this finding supports the bilingual advantage in inhibitory abilities, which should have reduced the *incongruity effect*, or in monitoring skills, which should have produced faster overall reaction times.

III. Experiment 6: Numerical Stroop Task in bilingual and monolingual children.

Experiment 5 was aimed at comparing the performance of large samples of bilingual children from a bilingual community to the large sample of matched monolingual children from monolingual environments, using the verbal Stroop task. However, it is worth keeping in mind that the nature of this task, based on word production, can distort the results and not reflect the real picture of bilinguals' and monolinguals' behavior. As it was mentioned in Chapter 2, many studies have consistently reported that bilinguals suffer a disadvantage in production of spoken responses (Ivanova & Costa, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005). For that reason, and following the same procedure that was applied in Chapter 2, Experiment 6 makes use of a less linguistically charged version of the same paradigm as the one used before –the numerical Stroop task (Besner & Coltheart, 1979; Kaufmann et al., 2005; Santens & Verguts, 2011).

Although attempts have been made to take a similar approach and use Stroop-like tasks in children to avoid the influence of linguistic variables (Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011), one could again observe that in these studies, relevant factors such as immigrant status are not reported or controlled for (although observing the uneven linguistic profiles of the participants one could infer that bilinguals had an important proportion of immigrants).Thus, the purpose of this experiment is to test whether the null results observed in children using the verbal Stroop task (Experiment 5) could also be replicated when the linguistic burden is (almost entirely) removed from the task demands. Furthermore, the aim is to compare these results to the ones obtained for the senior participants in verbal and numerical Stroop task (Chapter 2). This provides a great opportunity to investigate how inhibitory capacities develop over the course of schooling and to examine the extent to which those effects are modulated by the involvement of language (i.e., the bilingual advantage) in the tasks at stake. If bilinguals truly have better inhibitory skills than monolinguals, a reduced *conflict effect* should be observed in bilinguals as compared to monolinguals. Besides, as suggested by the general enhancement perspective (equated to monitoring by Costa and colleagues (2009) and denominated as the "BEPA" hypothesis by Hilchey and Klein in 2011), one could also predict global reaction time differences between the samples, with bilinguals being overall faster than monolinguals in the current tasks. However, considering the lack of consistency across the studies comparing monolingual and bilingual children, and taking into account recent evidence against the bilingual advantage in executive processing (see Paap & Greenberg, 2013; Paap, Johson & Sawi, 2015), unambiguous between-group differences should not be unquestionably expected.

Furthermore, considering that both Stroop tasks were performed by the same set of bilingual and monolingual participants, correlations between the indices will be analyzed in order to disentangle whether or not the same inhibitory processes are being applied in both tasks, which would be indicative of the cross-task reliability (Paap & Greenberg, 2013, Paap & Sawi, 2014).

1. Method

a. Participants

The participants for this experiment were the same ones as in Experiment 5 (see Table 8).

b. Materials

The materials used in this experiment corresponded to those used in Experiment 2.

c. Procedure

The experimental procedure was equal to the one used in Experiment 2.

2. Results

a. Latencies

Incorrect responses (less than 2.5% of the data) and reaction times below or above 2.5 standard deviations from the mean in each condition for each participant (less than 2.5% of the data) were excluded from the latency analysis. The mean latencies for correct responses and error rates are presented in Table 10. As in Experiment 5, different ANOVAs were conducted in order to explore the classic *Stroop effect*, the *incongruity effect* and the *congruency effect*, and their interaction with the linguistic profile. The same statistical approach to the one followed in Experiment 5 was used here.

Table 10.- Mean latencies for correct responses and error rates in all conditions for the numerical Stroop task for the groups of children in Experiment 6. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets. Effects are shown underneath the conditions.

		Reaction times				**Error rates**			
		Monolinguals		**Bilinguals**		**Monolinguals**		**Bilinguals**	
Conditions	**Congruent**	683	(194)	696	(206)	0.94	(2.5)	1.04	(2.76)
	Incongruent	727	(188)	737	(193)	4.54	(7.75)	3.82	(4.99)
	Neutral	695	(179)	703	(191)	0.89	(3.23)	1.24	(3.16)
	Total	701	(176)	712	(188)	2.12	(3.08)	2.03	(2.32)
Effects	**Stroop**	44	(80)	42	(80)	3.6	(8.14)	2.78	(5.58)
	Congruency	12	(67)	8	(55)	-0.05	(3.11)	0.2	(3.94)
	Incongruity	-32	(70)	-34	(76)	-3.65	(8.32)	-2.58	(5.68)

As in Experiment 5, an ANOVA including the factors Condition (3 levels: congruent, incongruent, neutral) and Language (2 levels: monolinguals, bilinguals) was run. The Condition effect was significant [$F(2,1004)=98.035$, $p<.01$], but the Language effect and the interaction between these factors were negligible [Fs<1 and ps>.52]. Next, the individual effects were explored[7].

ANOVAs conducted to explore the *Stroop effect* on the reaction times exploring the factors Condition (congruent vs. incongruent) and Language (monolinguals vs. bilinguals)

[7] As in the classic verbal Stroop, an additional analyses including the factor Grade was performed. *Stroop effect* analysis on RTs showed 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]. 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 [all ps.>.18]. In the analysis of the RT effects for the *incongruity* index 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] was found, but Grade [$F(5,492)=2.37$, p<.04] was significant. No effect of Language [$F(1,492)=2.85$, p=.09] and no interaction between them [F<1 and p>.73] was found either in error rates. The Grade effect demonstrated that the general *incongruity effect* augmented as a function of age for both language groups. RTs for congruency data showed no effect whatsoever [all Fs<1 and ps>.48]. For the net effects in the error rates, Grade was marginally significant [$F(5,492)=2.16$, p=.06], but no effect of Language [F<1, p>.42] or interaction [$F(5,492)=1.18$, p>.31] was encountered. The magnitude of the net *congruency effect* decreased with age.

showed a main effect of Condition [$F(1,502)=144.38$, $p<.01$], but no Language effect or interaction [all Fs<1 and ps>.48]. Participants took longer to respond to incongruent stimuli than to congruent stimuli (see Fig. 14). The *Stroop effects* were highly similar for the two Language groups, as supported by the Bayesian factor analysis (BF_{01}= 9.72).

The *incongruity effect* was analyzed by comparing the responses in the incongruent condition to those in the neutral condition. ANOVAs on the reaction times showed a significant effect of Condition [$F(1,502)= 102.70$, $p<.01$]. The Language effect and the interaction were not significant [all Fs<1 and ps>.33]. Longer reaction times were found for incongruent trials than for neutral trials (i.e., *the incongruity effect*). The *incongruity effect* was similar for the two Language groups, as indicated by the fact that Bayesian factor t-tests strongly supported the null hypothesis (BF_{01}= 9.72).

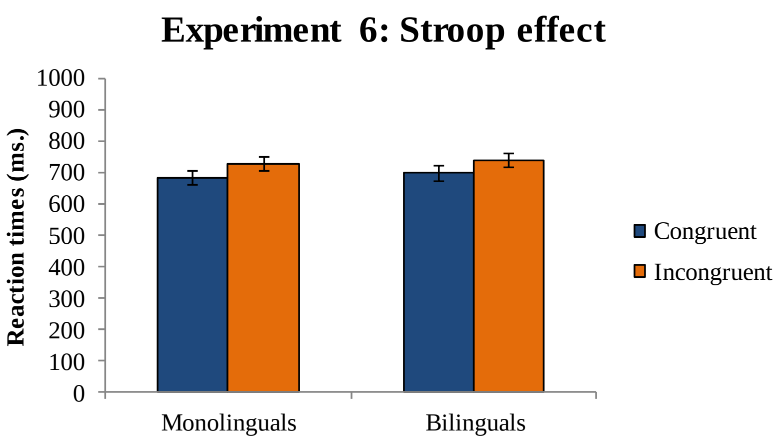

Figure 14. Reaction times for congruent and incongruent conditions in Experiment 6. Error bars represent confidence intervals of 95%

The ANOVAs on the reaction time data exploring the *congruency effect* (i.e., congruent vs. neutral conditions) showed a marginal main effect of Condition [$F(1,502)= 13.01$, $p<.01$], but no effect of Language or interaction [all Fs<1 and ps>.48]. Congruent trials were responded to faster than neutral trials. Bayes factor t-test analysis between groups favoured the null hypothesis (BF_{01}= 7.97).

b. Accuracy

The ANOVAs analyzing the *Stroop effect* (congruent vs. incongruent) including Language group (bilingual vs. monolingual) on the error data showed a main effect of Condition [$F(1,502)=144.38$, $p<.01$], and no effect of Language [$F<1$] or interaction [$F(1,502)=1.73$, $p>.18$]. Participants made more errors in the incongruent condition than in the congruent condition. The *Stroop effect* was similar for the two Language groups, as indicated by the Bayesian factor analysis (BF_{01}= 4.35).

Similarly, the analysis of variance exploring the *incongruity effect* (incongruent vs. neutral) on the error data showed a significant main Condition effect [$F(1,502)=96.19$, $p<.01$]. The Language effect was not significant [Fs<1 and ps>.31]. The interaction between these two factors was not significant [$F(1,502)=2.82$, $p<.1$]. The Condition effect was highly similar in both groups [monolinguals: $F(1,251)=48.34$, $p<.01$; bilinguals: $F(1,251)=52.02$, $p<.01$]. Bayesian factor analysis showed the tendency towards the null hypothesis as well (BF_{01}= 2.57).

Finally, when the error rates on the *congruency effect* (congruent vs. neutral words) were analyzed, Language was not significant [$F(1,502)=1.16$, $p>.28$]. The Condition effect and the interaction were also negligible [Fs<1 and ps>.22]. The *congruency effects* were virtually identical for the two Language groups as demonstrated by the Bayes factor t-test comparison (BF_{01}= 7.49).

Replicating what was found in Experiment 5, none of the indices obtained for the numerical Stroop effect was modulated by the linguistic profile of the participants, neither did they differ in overall reaction times. Importantly, the Bayesian null hypothesis testing favored the null hypothesis in the majority of the comparisons, revealing a very similar behavior of both groups in this task.

IV. Interim conclusions: Experiments 5 and 6

The results obtained in the verbal and numerical Stroop tasks replicated the patterns observed in seniors with the same tasks. In the reaction times, significant *Stroop effects* (difference between congruent and incongruent trials) are obtained, and they were similar across language groups. The same result was obtained when *congruency* and *incongruity effects* were calculated separately. The inclusion of other external factors in the analysis, such as grade, did not alter the non-significant effect of language. Error rates show a similar pattern. However, bilingual children showed a small but significantly larger *congruency* effect than their monolingual peers in the verbal Stroop task (see also Bialystok et al., 2004), which was probably partly due to differences in the baseline (the neutral condition). This difference cannot be explained by enhanced inhibitory abilities in bilingual children, because there is nothing that should be inhibited in the congruent condition. Similarly, it is not plausible that an enhancement of general monitoring abilities is driving this effect, since this enhancement would have been captured in all the other conditions as well as in the numerical Stroop task. The high number of participants tested and the reduced number of errors in these tasks, suggest that a significant difference in error rates in a single index should be interpreted with caution.

These results seem to go against previous results and indicate that the bilingual advantage in inhibition or monitoring (see Bialystok, Craik, Klein, & Viswanathan, 2004; Martin-Rhee & Bialystok, 2004; see Hilchey & Klein, 2011, for review) again fails to be replicated when the confounding factors are controlled for (see also Paap & Greenberg, 2013), as other authors previously showed in children (Gathercole et al., 2014).

As in the elderly sample, the cross-task coherence was analyzed by testing the correlation between the magnitude of each of the indices on both of the tasks, for both bilingual and monolingual children. All the correlations between the RT effects (*classic Stroop, congruency* and *incongruity* effects) were extremely mild (*classic Stroop*: r=.07; *congruency*: r=-.05; *incongruity:* r=.14), in spite of the large sample of participants being tested (N=504). This suggests that the generalizations based on effects that are not consistent across the indices and tasks would be weak, i.e., generalizations should not be made if there is little coherence between the indices of apparently similar tasks and paradigms (see also Miyake & Freedman, 2012, for a similar argument).

V. *Experiment 7: the Attentional Network Task in bilingual and monolingual children.*

In the previous two experiments I have shown that bilingual and monolingual children do not differ in their inhibitory or monitoring abilities when these are measured in the verbal Stroop task, or in the numerical Stroop task (which suggests that these results are not driven by the possible lexical access differences). However, it is worth noting that the numerical Stroop's results might not be completely free from other influencing factors such as semantics, since several studies have shown the impact of semantic features in number processing (Dehaene, 1992; Dehaene, Bossini, & Giraux, 1993; Fias, 1996; but see Fias, 2001 for results indicating asemantic written number word to phonetics translation). For that reason, the third experiment included in this chapter was conducted using the Attention Network Test (ANT; see Fan, McCandliss, Sommer, Raz, & Posner, 2002). This task, which is a combination of the classical flanker task (Eriksen & Eriksen, 1974) and the cueing task (Posner, 1980), measures the three independent attentional networks: *orienting, alerting* and *conflict* (e.g., Fan, Flombaum, McCandliss, Thomas, & Posner, 2003). In this task, participants need to respond to the presence of an arrow on the screen, by indicating whether the arrow is pointing to the left or to the right. The critical arrow (e.g., →) can be flanked by another 2 arrows on each side, either pointing in the same direction (congruent trials; e.g., →→→→→) or in the opposite direction (incongruent trials; e.g., ←←→←←). Simple lines can also flank the central arrow, this way creating the neutral condition (e.g., -- -- → -- --). The arrows can appear in either in the upper or in the lower part of the screen. Previous to each flanker trial and after a random time period, participants can be given cues about the position of the arrows, in the form of an asterisk. Thus, the Cue factor can be manipulated so that participants see a valid spatial cue (i.e., an asterisk in the same position as the upcoming arrows), a double cue (i.e., one asterisk in the upper part and another one in the lower part of the screen), a neutral cue (an asterisk in the middle of the screen) or no cue at all. With the combination of these 4 cue conditions (double, spatial, center and no cue) and 3 flanker conditions (congruent, incongruent and neutral), a measurement of the three attentional networks can be obtained. The index of the *alerting network* can be obtained by subtracting the reaction times in the double cue condition and the ones in the no cue condition. Similarly, the *orienting index* can be obtained by comparing the central cue and the spatial cue conditions. Finally, and possibly the most important for the purposes of this thesis, the *conflict index*, which is closely related to executive control, can be obtained by comparing the reaction times to incongruent and congruent trials.

In the Revised ANT task (ANT-R, Fan et al., 2009) a fifth cueing condition was created: the invalid spatial cue. This was conceived as the opposite of the valid spatial cue, where the asterisk precedes the target stimuli in the exact same position. The invalid spatial cue, on the other hand, precedes the target arrow in the opposite part of the screen, so that an asterisk in the lower part would precede targets appearing in the upper part of the screen, and an asterisk in the upper part would precede targets appearing in the lower part of the screen. By comparing the (longer) latencies in the invalid cue trials to the (shorter) latencies in the valid cue trials, the *validity index* is obtained, considered as an index of reorienting attention.

The ANT task has been found to show a different developmental pattern for the different networks. Rueda et al. (2004), tested children from 6 to 10 years of age in an adapted version of the ANT task where the arrows were replaced with fish to make it more child-friendly. Not surprisingly, they found that the overall reaction times and error rates decreased gradually as a function of age. When the *alerting, orienting* and *conflict* networks were analyzed separately, the authors found that the developmental pattern was not parallel for these three networks. On the one hand, the *alerting* network showed negligible changes between ages 6 and 10, while the *orienting* network failed to show a clear-cut developmental change. On the other hand, the *conflict* effect showed a remarkable improvement from age 6 to age 7, remaining relatively stable after that.

Similarly to the Stroop and the Simon tasks, an intriguing pattern has been found when the differences between bilinguals and monolinguals were explored using the ANT task. For instance, Costa and colleagues (Costa, Hernández, & Sebastián-Gallés, 2008) tested 100 young Catalan-Spanish bilinguals (mean age: 22 years) and compared them to 100 monolingual peers. Regarding the specific attentional networks, they found that monolinguals showed larger *conflict* effects than bilinguals. This was argued to reflect better inhibitory skills, in line with the bilingual advantage hypothesis. Nevertheless, this differential effect vanished during the course of the experiment, probably as a function of within-task specialization or adaptation strategies. Besides, in the a*lerting* network, bilingual participants benefited from the presence of an Alerting Cue more than monolinguals. They also reported that bilingual participants were overall faster than their monolingual peers, regardless of the Flanker and Cue type. The authors also showed that the overall RT differences could not be simply explained by bilinguals just being better than monolinguals at conflict resolution, given that they were also faster in congruent trials. Taken together, these results led the authors to abandon the hypothesis that the bilingual advantage was the consequence of bilinguals' better ability to process incongruent information, and to propose that it reflected bilinguals' enhanced monitoring abilities. This study (and other closely related findings, e.g., Bialystok & Martin, 2004; Carlson & Meltzoff, 2008) suggests that the bilinguals are overall faster in tasks that involve conflict resolution, and that the incongruity effect produced by the incongruent trials is larger for monolinguals than for bilinguals.

Upon comparing these results to previous literature, Costa and colleagues (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009) noticed that the magnitude of the *conflict effect*s and the overall RTs between groups were highly similar in a large portion of the studies reported in the literature (e.g., Bialystok et al., 2008; Bialystok, 2006), and that the amount of studies actually showing a bilingual advantage was rather limited (Costa et al., 2008; Martin-Rhee & Bialystok, 2008). In the cases where a bilingual advantage is found, it was most likely to be present in the form of an overall RTs difference between groups, rather than in the magnitude of the *incongruity* or *congruency effects* (Bialystok, Martin, & Viswanathan, 2005; Morton & Harper, 2007). To further test this hypothesis, Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, (2009) ran a version of the ANT with monolingual and bilingual participants, manipulating the monitoring demands . In the first experiment, they created a low-monitoring context, with 92% of the trials belonging to one condition (either congruent or incongruent) and 8% to the other condition, thus making the upcoming target highly predictable. In a second experiment, they created two high-monitoring contexts. In one of the contexts, each condition (i.e., congruent and incongruent) was represented by 50% of the trials, making it difficult to predict the condition of the upcoming trial. In the other context, the authors opted for a 75% congruent-25% incongruent distribution of the trials. Costa et al. found that bilingual participants were overall faster than monolinguals in the high monitoring contexts (namely, in the context with 50% of the trials per condition and, to a less extent, in the one with 75%-25%), but did not show differences in the magnitude of the *conflict index*. Contrarily, in the low-monitoring context, both groups behaved similarly, with no differences in overall RTs or in the magnitude of the *conflict effect*. In the 75%-25% context, an advantage was found in the overall RTs and in the *conflict effect* for bilinguals, but these effects were modest and confined to the first experimental block. Costa et al. argued that 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 the absence of it in the individual ANT indices (associated with the different components of the attentional network), was that the advantage stems from the conflict monitoring, rather than from the inhibitory capacity per se (Bialystok et al., 2004; Green, 1998; Kroll, Bobb, Misra, & Guo, 2008; and see also Morales, Gómez-Ariza, & Bajo, 2013 for an explanation combining inhibitory and monitoring skills). For this reason, the appearance of the bilingual advantage in inhibition seems to be restricted to certain experimental conditions, and is often not replicated (see, among others, Kousaie & Phillips, 2012; Paap & Greenberg, 2013; and Prior & MacWhinney, 2010; for null results in the flanker task).

According to the monitoring skills 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 explains the fact that the bilingual advantage in the overall RTs is found in mixed-design experiments rather than in block-design experiments. Namely, in the latter, there is no need for adaptation from one trial to the following one because participants can predict the within-block consistency. Still, the sample tested was composed of young adults, and, given the lack of consistency found in the studies testing children, the extent to which the bilingual advantage is present (or absent) in younger samples remains to be seen.

Clearly at odds with the findings reported by Costa et al. (2009), a study by Pelham & Abrams (2014) comparing young early bilinguals, late bilinguals and monolinguals in the ANT task, showed a significant bilingual advantage in conflict resolution. They found that monolinguals were slower than the two bilingual groups in incongruent trials, showing larger *conflict effects* than both late and early bilinguals (with no differences between the two bilingual groups).

Additionally, although the main focus of bilingualism research using the ANT task has been the *conflict* effect, (given its direct relationship with executive control and inhibitory skills), it is worth noting that there has also been a debate on the evidence for the differences in the *alerting* effect (Costa et al., 2008; but see Costa et al., 2009) and in the *orienting* network (Colzato et al., 2008; but see Hernández et al., 2010).

Hence, the objective of this experiment is to test whether the findings of the bilingual advantage using the ANT are reliable in children. If the differences between bilingual and monolingual children are found in the *conflict effect* and/or the overall RTs, this would invalidate the null results obtained previously with both Stroop tasks, indicating the possible influence of lexical access differences and providing a support to the bilingual advantage hypothesis. If, on the other hand, differences are not found in the *conflict network* (bilingual advantage in inhibition) or overall RTs (bilingual advantage in monitoring), it would be a further confirmation that the bilingualism is not enhancing executive functions.

1. Method

a. Participants

Two groups of participants, most of whom also took part in Experiments 5 and 6, were recruited from different schools in Spain (n=360, females=211). The first group was made up of 180 Spanish monolingual children (females=106) from second, third, fourth and fifth grades of elementary school and first grade of secondary school. These monolinguals were recruited from Spanish schools in places where Spanish is the only official language, and none of them had fluent knowledge of any other language than Spanish. Also, none of them belonged to any minority and they were only exposed to Spanish at home. The second group was formed of 180 bilingual children (females=105) from the same grades as monolinguals. They were all born and lived in the Basque Country, a Spanish region with two co-official languages - Basque and Spanish. All these bilingual children were attending bilingual schools where both languages were used as vehicular languages. According to the legal requirements, bilingual schools in the Basque Country ensure that the teachers switch from one language to the other as they switch academic subjects, making sure that a similar distribution of the languages across subjects and school time (50% in each language) is achieved. This way, Basque children attending bilingual schools are exposed actively to the two languages on a daily basis during schooling.

A linguistic competence questionnaire completed by 171 of the 180 bilingual children's parents (namely, 95% of the sample) showed that bilingual participants acquired the two languages very early in life, with overall age-of-acquisition scores of 0.58 years (SD=0.77) for Spanish and of 2.23 years (SD=1.07) for Basque. The parents' subjective ratings for the children's performance in Basque and Spanish were collected on a 0-to-10 scale, where 10 represented the perfect knowledge and use of language. Children's mean proficiency score in Spanish was 8.65 (SD= 1.17), and their score in Basque was 5.96 (SD= 1.63).

In order to explore the developmental trajectory of the attentional networks, the sample of bilinguals and monolinguals were divided into three evenly distributed subgroups. Monolingual and bilingual 2nd and 3rd graders were classified as Group 1, 4th and 5th graders were classified as Group 2, and 6th graders and students from the first grade of high school were classified as Group 3. 120 children were included in each group, half of them (n=60) corresponding to a monolingual language context and the other half corresponding to a bilingual context. Pairwise comparisons within each group showed no differences (all *p*s>.11) between bilinguals and monolinguals in age, gender, overall reading and arithmetic skills (as assessed by their teachers on a 1-to-5 Likert scale), verbal, non-verbal and composed IQ (obtained from the Spanish version of the Kaufman Brief Intelligence Test (1990), K-BIT), household income (classified according to the following categories: >3000€/month, category 1; 2001-3000€, category 2; 1601-2000€, category 3; 1201-1600€, category 4; 750-1200€, category 5 and <750€ category 6), number of years of formal education of the parents, and parental work status (including three possible categories: neither works, only one of them works, both of them work). Furthermore, none of the participants had any specific developmental, psychological, psychiatric or educational disorder, deficit or special need, as verified by including a series of questions in this regard in the questionnaires completed by parents and teachers. Besides, none of the children repeated any academic year and no child with scores below the 20th centile in verbal, non-verbal and combined IQ tests was included in the sample. Hence, the two groups were carefully matched in many socio-economic and cognitive measures (see Table 11 for detailed comparisons).

Table 11.- Characteristics of the bilingual and monolingual children tested in Experiment 7 divided by grade. Mean values are displayed with standard deviations between brackets.

		Age (in years)		Reading scores (1-5)		Math scores (1-5)		Verbal IQ (centiles)		Non-verbal IQ (centiles)		General IQ (centiles)		Incomes (category)		Parents' education (years)		Parents' work situation (category)	
Group 1	**Bilinguals**	7.57	(0.59)	4.53	(1.17)	4.52	(0.93)	77.18	(14.58)	63	(22.31)	68.82	(17.88)	1.98	(1.07)	14.3	(2.49)	1.9	(0.35)
Primary 2nd & 3rd	**Monolinguals**	7.55	(0.53)	4.57	(0.98)	4.57	(0.87)	79.28	(15.76)	60.85	(22.18)	69.73	(19.74)	2.15	(0.99)	13.88	(2.76)	1.9	(0.35)
	p value	***0.88***		***0.84***		***0.72***		***0.31***		***0.48***		***0.7***		***0.25***		***0.29***		***1***	
Group 2	**Bilinguals**	9.53	(0.57)	4.75	(0.95)	4.87	(0.89)	63.72	(18.62)	66.13	(18.43)	62.3	(17.56)	1.77	(0.96)	14.59	(2.16)	2	(0.)
Primary 4th & 5th	**Monolinguals**	9.5	(0.6)	4.78	(0.83)	4.82	(0.87)	65.32	(19.12)	66.53	(17.81)	63.32	(17.13)	1.88	(0.94)	14.44	(2.39)	2	(0.)
	p value	***0.78***		***0.84***		***0.75***		***0.65***		***0.9***		***0.76***		***0.55***		***0.71***		***1***	
Group 3	**Bilinguals**	11.4	(0.65)	4.57	(1.06)	4.42	(0.91)	56.93	(18.23)	68.03	(17.9)	59.52	(17.64)	1.48	(0.68)	14.62	(2.3)	1.92	(0.28)
Primary 6th & Secondary 1st	**Monolinguals**	11.5	(0.54)	4.58	(0.91)	4.63	(0.84)	61.2	(17.73)	63.1	(19.78)	59.37	(19.28)	1.65	(0.66)	14.07	(2.34)	1.95	(0.22)
	p value	***0.73***		***0.92***		***0.13***		***0.12***		***0.11***		***0.96***		***0.17***		***0.18***		***0.42***	
	Bilinguals	9.51	(1.69)	4.62	(1.06)	4.6	(0.93)	65.94	(19.11)	65.72	(19.64)	63.54	(18.02)	1.74	(0.93)	14.5	(2.31)	1.94	(0.26)
Total	**Monolinguals**	9.51	(1.7)	4.64	(0.91)	4.67	(0.86)	68.6	(19.14)	63.49	(20.03)	64.14	(19.14)	1.89	(0.89)	14.13	(2.5)	1.95	(0.24)
	p value	***0.93***		***0.79***		***0.42***		***0.16***		***0.31***		***0.77***		***0.13***		***0.12***		***0.66***	

b. Materials

In this version of the child Attention Network Test (ANT) two within-subject factors were manipulated, Cue type (double cue, valid cue, invalid cue, neutral cue and no cue) and Flanker type (incongruent, congruent), leading to a total of 10 conditions. As already explained in the introductory section of the current experiment, Fan et al. (2009) suggested that the inclusion of an index of *validity* within the cueing conditions provides an additional measure of the ability to reorient attention. Hence, valid and invalid cues were included in the current design too. The Cue manipulations were created by presenting (or not) an asterisk on the screen prior to the presentation of the target strings. These cues could be presented at the same position of the upcoming target (valid condition), or in the opposite position (invalid condition). In order to create the double cue condition, two asterisks were presented at the same time above and below the center of the screen. The neutral cueing condition was created by presenting the asterisk at the center of the screen, and the no cue condition was created by not providing any visual cue. Regarding the Flanker manipulation, the target was a left- or right-pointing yellow fish (1.6°), presented above or below the fixation cross. This central fish was flanked on both sides by two fish pointing either in the same direction (congruent trials), or in the opposite direction (incongruent trials). The distance between the fish was 0.21°. The target and flankers subtended 8.84° and were presented 1° above and below the fixation cross over a blue-green background. For detailed description of the stimuli and procedure, see Rueda et al. (2004).

c. Procedure

All the stimuli were presented on a computer screen. Each trial began with a fixation cross (1° of visual angle) with a random duration between 400 and 1600ms. Then a cue (an asterisk) would appear in any of its variants for 150ms. Next, a centered fixation cross appeared on the screen for 450ms, immediately followed by the target and flanker stimuli. The target string stayed on the screen until a response was given or for a maximum of 1700ms. After each trial, feedback was provided.

A session of the ANT consisted in a total of 288 trials. Each trial represented one of the 10 conditions mentioned above (Cue type x Flanker type). To keep the high-monitoring demanding context (see the introduction to this experiment), 50% of the trials belonged to the congruent condition and the other 50% to the incongruent condition. Regarding each cueing condition, there were 72 double cue, 48 valid, 48 invalid, 48 neutral cue and 72 no cue trials. Participants were seated at a distance of about 55cm from the screen and they were instructed with a series of practice trials to indicate the direction of the central fish in the strings, by pressing the "L" key in the keyboard to indicate right or the "S" key to indicate left. Both accuracy and reaction times were recorded on each experimental trial.

2. Results

Reaction times below 200ms (only representing 0.12% of the data) were excluded. Reaction time data was trimmed using the classic 2.5 SD criterion, resulting in the exclusion of the 2.49% of the data. The RTs associated with erroneous responses were not included in the latency analyses. Before focusing on the individual indices for each attention network, all the conditions were analyzed in a general 5 x 2 x 2 ANOVA including Cue Type (no cue, valid cue, invalid cue, double cue and neutral cue) and Flanker Type (congruent and incongruent) as within-participant factors, and Language (bilinguals and monolinguals) and Group (first, second and third group) as between-participants factors. In subsequent analyses the different attention networks were explored by measuring the following indices: the difference between congruent and incongruent trials as a reflection of inhibitory control (*conflict effect*), the differences between the double cue and the no cue conditions for the alerting network (*alerting effect*), the *orienting* network as measured by the difference between the neutral cue and valid cue trials (*orienting effect*), and finally the difference between valid cue and invalid cue trials as markers of the *validity effect*. Detailed information about the RT and error data is presented in Table 12. Furthermore, as the classical hypothesis testing does not allow for accepting the null hypothesis, the critical differences of interest were also tested following the Bayesian approach (see Rouder et al. 2009, among others). For each index (*conflict, validity, orienting* and *alerting*), the Bayes factor (BF) approach was used to compare a model that assumed no differences between bilinguals and monolinguals (H_0) with a model that assumed that bilinguals perform differently from monolinguals (H_1).

a. General analysis

In the RT analysis, significant main effects of Flanker Type [$F(1,354)= 1624.68$, $p<.01$], Cue Type [$F(4,1416)= 237.19$, $p<.01$] and Group [$F(2,354)=120.07$, $p<.01$] were found. In contrast, the main effect of Language was not significant [$F(1,354)=2.22$, $p>.13$]. The two-way interaction between Flanker Type and Group was significant [$F(2,354)=12.5$, $p<.01$], and the same was true for the interaction between Flanker Type and Cue Type [$F(4,1416)=24.12$, $p<.01$]. None of the other interactions was significant.

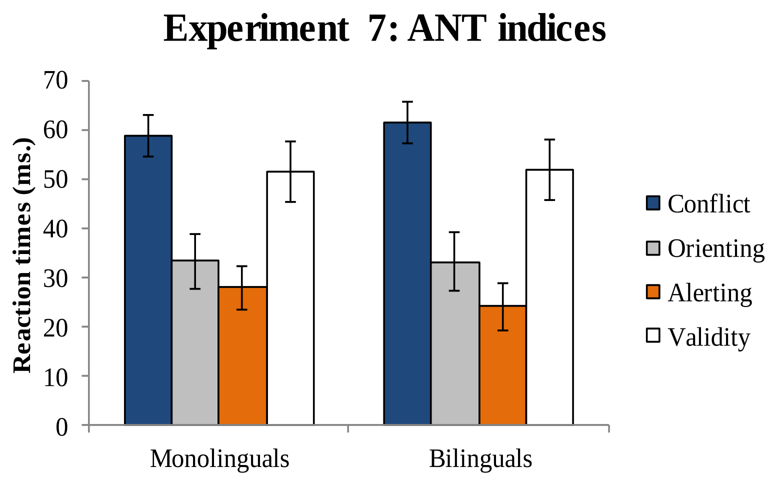

Figure 15. Indices obtained for each language group in Experiment 7. Error bars represent confidence intervals of 95%

In error rate analysis, both Language groups performed similarly ($F<1$). The main effects of Flanker Type [$F(1,354)=303.20$, $p<.01$], Cue Type [$F(4,1416)=11.52$, $p<.01$], and Group [$F(2,354)=43.53$, $p<.01$] were significant. The only significant interactions found were the Flanker Type X Group interaction [$F(2,354)=6.85$, $p<.01$], and the Flanker Type X Cue Type interaction [$F(4,1416)=90.32$, $p<.01$].

Thus, it is important to notice that none of the interactions with Language was significant, showing that the same effects hold for bilinguals and monolinguals.

Considering the reliable Flanker Type X Cue Type interactions, and following preceding research, each of the effects mentioned above were explored individually (i.e., *conflict, alerting, orienting* and *validity*), and the manner in which the between-participants factors Group and Language could modulate them (see Table 12 and Fig. 15 for comparisons between Language groups for each index, and see Table 13 for a detailed comparison between Language Groups in each Age Group).

Table 12.- Mean latencies for correct responses and error rates in all conditions for the ANT task for the groups of children in Experiment 7. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets. Attentional network indices are shown underneath the conditions.

		Reaction times				Error rates			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	Double Cue	676	(101)	690	(110)	4.58	(5.64)	4.92	(5.4)
	Neutral Cue	693	(111)	706	(113)	4.99	(5.72)	4.92	(5.41)
	Valid Cue	660	(107)	673	(107)	4.2	(5.84)	4.56	(5.3)
	Invalid Cue	711	(109)	724	(103)	5.6	(6.43)	5.91	(6.3)
	No Cue	704	(105)	714	(108)	5.02	(5.68)	5.69	(5.61)
	Congruent	659	(104)	671	(104)	3.18	(4.78)	3.45	(4.58)
	Incongruent	718	(107)	732	(109)	6.57	(6.3)	6.95	(5.9)
	Total	689	(104)	702	(106)	4.88	(5.28)	5.2	(4.91)
Networks	Conflict index	59	(29)	61	(30)	3.39	(3.71)	3.5	(3.91)
	Orienting index	33	(39)	33	(40)	0.79	(3.77)	0.36	(4.22)
	Alerting index	28	(31)	24	(33)	0.43	(4.09)	0.76	(4.)
	Validity index	52	(43)	52	(42)	1.4	(4.28)	1.35	(4.51)

b. Executive network: the conflict effect

In the RT analysis, the *conflict effect*, as measured by the factor Condition (congruent vs incongruent trials), was significant [$F(1,354)=1624.68$, $p<.01$], as well as the main effect of Group [$F(2,354)=120.07$, $p<.01$], and the interaction between them [$F(2,354)=12.50$, $p<.01$]. It took longer for participants to respond to the incongruent trials as compared to the congruent ones, and participant speed of response increased as a function of age (see below). Importantly, the main effect of Language was not significant [$F(1,354)=2.22$, $p>.13$], and it did not interact with Condition ($F<1$) or with Group ($F<1$). The three-way Language X Condition X Group interaction was not significant [$F(2,354)=2.22$, $p>.11$]. Hence, it can be concluded that

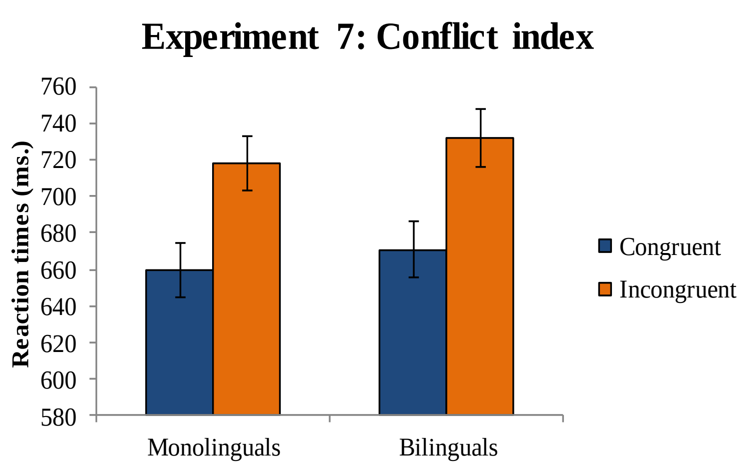

Figure 16. Reaction times for congruent and incongruent conditions in Experiment 7. Error bars represent confidence intervals of 95%

monolinguals and bilinguals showed highly similar *conflict effects* (see Fig. 16). This conclusion was strongly supported by Bayesian t-test comparisons (BF_{01}=5.94).

In order to explore the origin of the significant Condition X Group interaction, follow-up contrasts were run collapsing the data across linguistic profiles. Pairwise contrasts showed that the differences in the responses to the two types of Flankers (congruent, incongruent) decreased with age. Thus, when comparing the *conflict effect* in each Group, it can be observed that the first group showed the largest *conflict effect* (average of 70ms), and that this effect progressively diminished with age (Group 2= 57ms; Group 3= 52ms). Pairwise tests showed that the effect was significantly larger for Group 1 relative to Group 2 and Group 3 (Group 1 vs. Group 2: $t(238)=3.18$, $p<.01$; Group 1 vs. Group 3: $t(238)=4.54$, $p<.01$), while the difference was not significant between Groups 2 and 3 ($t(238)=1.70$, $p>.1$).

In error rate analysis, only the main effects of Condition [$F(1,354)=303.20$, $p<.01$] and Group [$F(2,354)=43.53$, $p<.01$] were significant. The only significant interaction was found between Condition and Group [$F(1,354)=6.85$, $p<.01$]. Replicating the RT data, the error data showed a clear *conflict effect*, with higher error rates in incongruent than in congruent conditions and a modulation of the percentages of errors as a function of age (i.e., overall error rates diminished as a function of age). Given the significant interaction, it can be concluded that the magnitude of the *conflict effect* decreased as a function of age. Importantly, the Language effect and the interactions between this and the other factors were negligible (all Fs<1 and all ps>.5). As in the reaction times, Bayesian t-test analysis fully supported the alternative hypothesis over the null (BF_{01}=8.26).

c. Alerting network: the alerting effect.

Considering the differences in RTs between the double cue and the no cue conditions, only the main effects of Condition [$F(1,354)=239.44$, $p<.01$] and Group [$F(2,354)=118.55$, $p<.01$] were significant. The Language effect was not significant [$F(1,354)=2.05$, $p>.15$]. None of the interactions was significant (Fs<1.20, ps>.27). Hence, participants responded faster to double cue trials than to no cue trials and they became overall faster as their age increased, but the difference between the cueing conditions did not differ across ages or across language profiles. Comparison between language groups performed by Bayesian t-test analysis showed that the null hypothesis was the best fit for the data (BF_{01}=4.82).

In the error rate analysis, the only significant effects corresponded to the factors Condition [$F(1,354)=7.81$, $p<.01$] and Group [$F(2,354)=41.25$, $p<.01$], showing that participants made more errors in no cue trials than in double cue trials and that the number of errors decreased as a function of age. No other effects or interactions were significant (all Fs<1.1 and all ps>.3). Bayesian analysis fully replicated what the analysis of latencies reported, supporting the null hypothesis ($BF_{01}=6.42$).

d. Orienting network: the orienting effect.

The *orienting* index (i.e., valid cue vs. neutral cue) was significant [$F(1,354)=260.30$, $p<.01$], as was the main effect of Group [$F(2,354)=109.45$, $p<.01$]. Responses to trials with a valid cue were faster than responses to trials with a neutral cue, and average RTs decreased as a function of age. In contrast, the main effect of Language was not significant [$F(1,354)=2.12$, $p>.14$], and none of the interactions involving the factor Language was significant (all Fs.<1). The null differences were again supported by the Bayesian t-test analysis ($BF_{01}= 8.59$).

A marginal interaction between Condition and Group was found [$F(2,354)=2.84$, $p<.07$], suggesting that the magnitude of the *orienting* effect decreased with age. Follow-up pairwise contrasts showed similar *orienting* effects for Groups 1 and 2 (39ms and 34ms, respectively; $t<1$), and a significantly smaller effect for Group 3 (27ms; Group 1 vs. Group 3: $t(238)=2.32$, $p<.03$; Group 2 vs. Group 3: $t(238)=1.71$, $p<.09$).

In the error rate analysis, the only significant effects were in Condition [$F(1,354)=7.33$, $p<.01$], showing more errors in the neutral cue than in the valid cue condition, and Group [$F(2,354)=34.74$, $p<.01$], showing a decrease in the amount of errors as a function of age. No other effects or interactions were significant (all Fs<1.1 and all ps>.3). Again, no differences were evident when language groups where compared with Bayesian comparisons ($BF_{01}= 5.23$).

e. Reorienting: the validity effect.

The difference between the valid cue and invalid cue trials was significant in the RT analysis [main Condition effect: $F(1,354)=539.92$, $p<.01$].The Group effect was also significant [$F(2,354)=117.92$, $p<.01$]. Invalid cues produced longer response times than valid cues, and the overall response times decreased as a function of age. These two factors (Cue and Age) marginally interacted with each other [$F(2,354)=2.78$, $p<.07$], suggesting that the magnitude of the *validity* effect decreased with age. Follow-up t-tests showed that the magnitude of the *validity* effect was similar for Groups 1 and 2 (54ms and 56ms, respectively; $t<1$). However, the effect was smaller for Group 3 (44ms) than for Group 2 ($t(238)=2.44$, $p<.02$) and, although marginally, than for Group 1 ($t(238)=1.84$, $p<.07$). Critically, the main effect of Language was not significant [$F(1,354)=2.37$, $p>.12$], and none of the interactions involving the Language factor were significant either (all Fs<1.15 and ps>.32). The hypothesis of no-differences was the best fit for the comparisons between language groups ($BF_{01}=8.58$).

Parallel findings were also observed in the error rate analysis, showing significant Condition [$F(1,354)=35.60$, $p<.01$] and Group effects ($F(2,354)=37.15$, $p<.01$), together with a marginal interaction between these two factors ($F(2,354)=3.03$, $p<.06$). Again, no other effects or interactions were significant (all Fs<1 and all ps>.5). There were no differences between language groups on the *validity* index ($BF_{01}= 8.54$).

Table 13.- Mean indices obtained for each age group in the ANT task, divided bilinguals and monolinguals. Mean effects are showed in milliseconds with standard deviations between brackets.

	Conflict effect				Orienting effect				Alerting effect				Validity effect			
	Bilinguals		Monolinguals		Bilinguals		Monolinguals		Bilinguals		Monolinguals		Bilinguals		Monolinguals	
Group 1	74	(36)	67	(36)	38	(50)	40	(50)	22	(41)	30	(41)	52	(50)	57	(48)
Group 2	54	(22)	61	(26)	34	(38)	34	(36)	24	(33)	25	(21)	54	(43)	58	(44)
Group 3	56	(25)	49	(20)	27	(29)	26	(26)	27	(22)	28	(27)	49	(31)	39	(34)

Thus, none of the classic indices obtained in the ANT task was modulated by the linguistic profile of the participants, in either latencies or error rates. Importantly, the Bayes Factors strongly supported the null hypothesis as the best fit for the data for every index, indicating a similar performance of both bilinguals and monolinguals in all of the tests.

VI. *Interim conclusions: Experiment 7*

The objective of this experiment was to test whether bilinguals show an enhancement in inhibitory control as compared to monolinguals in the ANT task (e.g., Kapa & Colombo, 2013; Pelham & Abrams, 2014), whether the potential bilingual advantage in monitoring would show up as an overall RTs decrease in responses (e.g., Costa et al., 2009), or, once the participant samples are properly matched and the external factors are controlled for, both groups would behave similarly (Paap & Greenberg, 2013). None of the indices explored with the present task was modulated by the linguistic profile of the participants, and their general RTs (irrespective of conditions) did not differ depending on whether they were bilinguals or monolinguals.

It is worth noting that, as in the past literature, the task showed robust and strong c*onflict, alerting, orienting* and *validity* effects (e.g.,Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Fan & Posner, 2004; Ishigami & Klein, 2010; Mackie, Van Dam, & Fan, 2013; Wang & Fan, 2007; Yin et al., 2012 among many others). Similarly, the developmental pattern obtained was similar to the one observed recently by Rueda et al. (2004) in monolinguals. Both their and the current study show that age significantly modulates *conflict effect*, slightly modulates *validity* and *orienting* effects, and does not modulate the *alerting* index.

The fact that previous findings in indices and developmental trends fully replicate in the present study leads me to understand that the lack of differences between bilinguals and monolinguals is not a product of poor statistical power, but indeed the null hypothesis appears to be the best explanation for the presented data. In a nutshell, in spite of the statistical power of the current study, no significant differences between bilingual and monolingual children emerged in any of the tasks. Furthermore, using the Bayesian approach, the null hypothesis is undoubtedly the strongest candidate.

These results (see also Gathercole et al., 2014) are important for the debate regarding the bilingual advantage in attention skills, as they suggest that the alleged advantage may well be the result of uncontrolled factors and experimental design (e.g., Morton & Harper, 2007; Paap & Greenberg, 2013; see also Paap & Liu, 2014 and Paap, Johnson, & Sawi, 2015; Costa et al., 2009).

VII. General discussion: Bilingual and monolingual children

In the studies described in this chapter, I aimed at exploring the influence of the linguistic profile of children (monolingual or bilingual) on their executive functions in a large sample of monolingual and bilingual children of different ages, using two kinds of Stroop tasks and the ANT paradigm. The purpose was to establish whether bilingual children, as compared to carefully matched monolingual peers, exhibit enhanced EF skills in a linguistic and a non-linguistic task, under the assumption that this might arise due to their daily use of (and switching between) two languages. Overall, the pattern of results obtained does not confidently and reliably allow for such a strong conclusion, given that the evidence favouring a difference between monolinguals and bilinguals in inhibitory skills was not found. On the contrary, both the classic ANOVAs and Bayes factor analysis supported the null hypothesis of no differences between language groups.

The reaction time patterns observed in the classic and numerical versions of the Stroop task (Experiments 5 and 6, 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 (see Qiu, Luo, Wang, Zhang, & Zhang, 2006; for review). Therefore, other measures had to be considered in order to explore potential differences between groups, investigating separately *congruency* and *incongruity 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, a parallel set of analyses was performed to explore 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, no bilingual advantage was observed in the reaction time data. According to recent evidence on the bilingual advantage that suggests enhanced inhibitory skills in bilinguals (see Bialystok, Craik, Klein, & Viswanathan, 2004; Martin-Rhee & Bialystok, 2004; see Hilchey & Klein, 2011, for review), one would predict a diminished *incongruity* effect in bilingual compared to monolingual children. In contrast to this prediction, the data suggest that bilinguals performed similarly to their monolingual peers. Importantly, I analyzed 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, and again, I failed to find any significant effects of the linguistic profile of the participants (nor an interaction between grade and linguistic profile).

Regarding the effects observed in the error rates, the pattern very much resembles the one obtained in the RT analyses, opposing the idea of enhanced cognitive control in bilingual children. The only difference between monolinguals and bilinguals in the error data was found in the classic Stroop task (Experiment 5). In this task, there was no supporting evidence for group differences in the *Stroop* or *incongruity effects*. 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 direct link between this effect and any sort of enhancement in inhibitory skills in bilinguals, given that the difference in congruent trials is caused by differences in the baselines (with no differences in the incongruent trials). However, this small difference in the *congruency effect* could be potentially accommodated within the theories that posit the locus of the bilingual advantage in the general executive functioning level, rather than in the concrete level of inhibitory mechanisms (e.g., Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; Costa et al., 2009). Nonetheless, this difference was not replicated in the error data of the numerical Stroop task (Experiment 6). Given the high number of participants tested, the reduced number of errors in these tasks and the lack of cross-experiment replicability of this effect, the degree of generalization based on this error effect is highly limited and great caution is advised in this regard.

The data show unequivocal *incongruity* and *congruency effects* all across the range of ages in both RTs and accuracy. In both Stroop tasks, congruent trials were responded to faster than neutral trials (i.e., the *congruency effect*), and incongruent trials were responded to slower than neutral ones (i.e., the *incongruity effect*). Furthermore, overall RTs decrease as a function of increasing age, thus the older participants were faster than the younger ones. However, the lack of interactions between the magnitudes of these indices (*Stroop* and *incongruity effects*), which are classically interpreted as indicators of inhibitory control, and the linguistic profiles of the participants (monolinguals vs. bilinguals) contradicted the existence of bilingual advantage in inhibitory skills. Also, the absence of absolute differences between linguistic groups also suggested that overall RTs were similar between monolinguals and bilinguals, opposing the existence of enhanced monitoring abilities.

Interestingly, remarkably low correlation was found between each of the indices of the two Stroop tasks. Given the scarce coherence and consistency of similar indices across similar tasks and paradigms (see also Miyake & Freedman, 2012, for a similar argument), it is difficult to generalize arguments supporting the bilingual advantage based on inhibitory skills.

The third experiment aimed at investigating whether bilingual children exhibit an advantage as compared to their monolingual peers, using the ANT task (Experiment 7), which has been typically considered as the paradigm best suited to explore the different attention networks. Considering the existing debate between researchers suggesting that bilinguals outperform monolinguals in the ANT task (e.g., Kapa & Colombo, 2013; Pelham & Abrams, 2014) and those suggesting that the bilingual advantage in this task is restricted to certain conditions and designs (e.g., Costa et al., 2009), the question under scope here was whether a large sample of bilingual children would exhibit better performance in this task than a group of carefully matched monolingual children. Results unambiguously demonstrated that the so-called bilingual advantage could not be replicated in the ANT task when a sufficiently large and well-matched group of bilingual and monolingual children were tested.

It is worth mentioning that the lack of a bilingual advantage in this study cannot be ascribed to a general lack of sensitivity of the design to the specific attention network(s) that may underlie bilinguals' and monolinguals' performance. Replicating previous evidence from the monolingual domain, bilingual and monolingual children exhibited longer latencies and higher error rates for incongruent trials than for congruent trials (namely, a significant *conflict effect*). Similarly, both groups showed better performance in the double cue trials as compared to the no cue trials (namely, a significant a*lerting effect*). Also, participants' responses to the valid cue trials were faster and more accurate than their responses to central cue (i.e., a significant *orienting effect*). Finally, participants showed longer RTs and higher error rates in trials involving an invalid cue than in trials with a valid cue (i.e., a significant *validity* effect). Hence, considering that the current results fully replicate the indices observed in earlier studies with the ANT task (e.g., Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Fan & Posner, 2004; Ishigami & Klein, 2010; Mackie, Van Dam, & Fan, 2013; Wang & Fan, 2007; Yin et al., 2012 among many others), it is hardly possible that potential differences between bilinguals and monolinguals were masked due to a lack of statistical power of the current study (see also the magnitude of the F-values in this regard). Furthermore, from a developmental point of view, the current study has replicated and extended the findings observed by Rueda et al. (2004), where a smaller group of monolingual children was tested. The same developmental trend observed in that study can be seen here, suggesting that the *conflict effect* (hence, reflecting inhibition abilities) is the attentional index that is most sensitive to a developmental change, greatly decreasing as a function of age. On the other hand, more modest changes are seen in the v*alidity* and o*rienting effects* (note that the interactions were marginally significant in spite of the sample size), and no significant changes in the a*lerting effect* as a consequence of age.

In summary (see Table 14), no significant differences between bilingual and monolingual children emerged in their performance in none of the task and, importantly, Bayesian analysis strongly supports the null hypothesis as the strongest candidate. Hence, the results presented in this chapter add to a growing body of evidence showing that previous evidence of bilingual advantage may have emerged due to uncontrolled factors (e.g., Morton & Harper, 2007; Paap & Greenberg, 2013; see also Paap & Liu, 2014 and Paap, Johnson, & Sawi, 2015) or specific conditions associated with the design and procedure (e.g., Costa et al., 2009). The results presented in this Chapter, together with recently published articles (Gathercole et al., 2014) defend that bilingual and monolingual children behave comparably in tasks that tap into executive function abilities.

Table 14.- Update and summary of the results obtained in the present chapter. A cross indicates non-significant effects of language in the measures obtained. Correlations are reported only if significant.

	Tasks used	Inhibition	Monitoring	Between-task correlation
Elderly	*Verbal Stroop*	✕	✕	Incongruity in Exp. 3-4, r= -.52 Rest, n.s.
	Numerical Stroop	✕	✕	
Children	*Verbal Stroop*	✕	✕	N.s.
	Numerical Stroop	✕	✕	
	ANT	✕	✕	

Certainly, the observed lack of bilingual advantage should not be generalized to other age groups, given that the claims stated here are restricted to the conclusion that the so-called bilingual advantage in tasks focusing on participants' executive function skills is inexistent, or at best, extremely inconsistent and elusive in children. The aim in the next chapter is to test large samples of young adults in the main tasks that have been used to explore the bilingual advantage in inhibition and monitoring, in order to see whether the absence of an advantage observed in Chapters 2 and 3 is generalizable to adult samples and to other tasks.

Chapter 4: The bilingual advantage hypothesis in young adults

I. Overview and theoretical introduction

As explained in Chapter 1, even the defenders of the bilingual advantage have argued that "there is thus some evidence that the bilingual advantage is greatest in children and in older adults, but less constantly present in young adults – perhaps because the young adult group is at the developmental peak age for cognitive control" (Bialystok, Craik, & Luk, 2012, pp. 5-6). However, the only conclusions that can be driven with certainty from the two previous chapters of the present thesis is that no bilingual advantage arises from carefully matched large samples of bilingual and monolingual children and seniors. The results I presented in those chapters unambiguously indicate that there is no difference in the performance between monolinguals and bilinguals when inhibitory or monitoring capacities where compared. However, it still cannot be concluded that the bilingual advantage in lifelong balanced bilinguals is a delusion, as results from other age groups are necessary to compliment the current data. Young adults constitute the population which has been most extensively studied in the bilingual advantage studies, as it is the case with the majority of psycholinguistics research. Therefore, the effects of bilingualism on executive functions in young adult samples need to be explored to be able to derive strong conclusions regarding this issue. However, as it was stated in Chapter 1, the potential bilingual advantage effects are most elusive in this population, due to the cognitive peak that is being experienced at this age span. In addition, since the bilingual advantage has been investigated using several tasks that have not been used in the previous experiments with the elderly and children, instead of using only two classic tasks, several other tasks will be used to investigate any potential differences in EF between bilinguals and monolinguals. This is especially important considering the low cross-task replicability reported in the previous chapters, as well as in the literature (Paap & Greenberg, 2013, Paap & Sawi, 2014).

1. Previous evidence on the bilingual advantage in young adults

Results reporting an advantageous performance of bilinguals over monolinguals peers in the samples of young adults are not few. For example, Bialystok (2006) found that the young bilinguals showed speeded responses on the incongruent trials in the Simon task, which is a condition that requires conflict resolution. This was later replicated with young adults and seniors, although the bilingual advantage was found to be greater in the elderly sample (Bialystok, Craik, Klein, & Viswanathan, 2004). Subsequently, Bialystok and colleagues (Bialystok, Craik, Grady, Chau, Ishii, Gunji, & Pantev, 2003) failed in finding behavioral differences, but different brain patterns for adult bilinguals and monolinguals were observed when responding to the Simon task using the MEG technique (other studies also failed in finding a Simon task advantage in young adults, e.g. Bialystok, Martin, & Viswanathan, 2005). Moreover, Bialystok, Craik and Luk (2008) tested young and old adults in a modified version of the Simon task and in the verbal Stroop task, and found that bilinguals were less harmed by the incongruent trials than their monolingual peers in both tasks. Similarly, Bialystok and DePape (2009) found that bilinguals (and musicians) outperformed monolinguals in the Simon task. When it comes to the flanker task, similar results were reported: Abutalebi et al. (2012) showed a more efficient use of the anterior cinculate cortex (ACC, related to domain-general executive control) of bilinguals who outperformed monolinguals in the flanker task. This is in line with the findings by Marzecová et al., (Marzecová, Asanowicz, Krivá, & Wodniecka, 2013), who showed that bilinguals are less affected by the *conflict cost* than the monolinguals (see also Kapa & Colombo, 2013; Pelham & Abrams, 2014). Tao and colleagues (Tao, Marzecova, Taft, Asanowic, & Wodniecka, 2011) compared monolinguals, early bilinguals and late bilinguals in the lateralized version of the ANT task, and even though they reported a general bilingual advantage, the late bilinguals (who were also more balanced in proficiency) showed the best performance in conflicting conditions.

It is important to mention again that the studies showing bilingual advantage are based on small sample sizes, which leads to small power and limited strength in the interpretation of the results (Paap, Johnson, & Sawi, 2014). Furthermore, various confounding factors are not controlled for in many of these works. As one of the most obvious examples, Tao and colleagues' (2011) experiment presents data from bilinguals that have higher IQ scores (as measured by the Raven Advanced Matrices, see Raven, Raven, & Court, 2004) and a lower

SES than their monolingual counterparts. The authors tried to account for these differences by including those factors as covariates, which arguably ought to be avoided, considering that the covariates and the grouping factors should be independent (Miller & Chapman, 2001, see also Marzecová et al., 2013, for the same problem).

It is worth noting, that the studies showing null differences between bilingual and monolingual young adults are not scarce. For example, Paap & Sawi (2014) tested nearly 60 bilinguals and 60 monolinguals in the Simon and the ANT tasks, as well as the antisaccade and the colour-shape Stroop task. They found no evidence for bilingual advantage, as the group differences were non-significant. Moreover when the group differences were significant, they indicated a *monolingual* advantage. Kousaie and Phillips (2012a) also found no differences between young adult bilinguals and monolinguals in the Simon, Flanker and Stroop task. Paap & Greenberg (2013) found no bilingual advantage in flanker and Simon tasks. Therefore, it can be easily seen that the body of evidence suggesting that bilingual advantage in young adults is actually nonexistent or, at best, restricted to very specific settings, is growing (see Paap & Greenberg, 2013, Paap, Johnson, & Sawi, 2014; Paap, Johnson, & Sawi, 2015a;b; for review)

2. *Aim of the chapter*

The current chapter aims at exploring the reliability and replicability of the previous findings reporting the benefits of bilingualism on executive functioning, by testing large cohorts of carefully matched monolingual and bilingual young adults. In order to account for methodological issues that were raised by the studies that oppose the bilingual advantage proposal (Duñabeitia & Carreiras, 2015), this chapter analyzes the data of large young adult samples, adequately matched, in the same tasks that were used in the previous studies showing a bilingual advantage. Ninety young bilingual adults from the Basque Country (a region of the north of Spain where Basque and Spanish are co-official) and 90 carefully matched monolinguals from Murcia (a south-eastern region of Spain where only Spanish is spoken and official) were tested. This way, the historical, cultural and social backgrounds were shared among all of them, as inhabitants of the same country, but language exposure was critically different: while bilinguals have been immersed in a bilingual society where two languages are co-official and present in every different aspect of their lives, monolinguals live entirely in a monolingual society. As in the previous experiments, the samples were carefully selected, matching the monolinguals and bilinguals on all the factors that have been suggested

(Morton & Harper, 2007; Hilchey & Klein, 2011; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015a;b) to influence results in the studies of this type, such as SES, ethnicity or immigrant status (Mezzacappa, 2004; Noble, Norman, & Farah, 2005).

Importantly, the cross-tasks replicability in this field is low, and the tasks used to measure the bilingual advantage rarely correlate (as it has been shown in the previous Chapters, see also Paap and Greenberg, 2013). Thus, it is important to use several different tasks to try to capture any potential differences that are rather elusive in the sample tested here (i.e., young adults). Hence, participants will be tested in all the main classic tasks traditionally used to examine the bilingual advantage in inhibitory control and monitoring. Namely, participants will be tested in the flanker (Eriksen & Eriksen, 1974), Simon (Simon & Rudell, 1967) and two variants of the Stroop (Stroop, 1935) tasks. Thus, young adults are going to be tested in the same tasks as the elderly (i.e., both Stroop tasks) and children (i.e., both Stroop and the flanker task, which is in essence the source of the *conflict effect* of the ANT). The Simon task was added given its extensive use in the bilingual advantage literature, in order to be able to extract more objective information and have a clearer picture. The commonality of these three tasks, as explained in previous sections, is that they all include trials where every piece of information presented favours the target response (i.e., congruent trials) and trials where some strong and salient information favours the opposite response to the one that needs to be produced (i.e., incongruent trials). The particularities of each task will be explained within each individual experimental section.

As explained in Chapter 1, in all these tasks, the difference between the congruent and incongruent trials (the *conflict effect*, or the *Stroop effect* in the Stroop task) has been taken as an indicator of the inhibitory skills. If the bilinguals show reduced differences, this could be ascribed to their better ability to deal with incongruent and conflicting information (the bilingual advantage in inhibition, or the “BICA” perspective, Hilchey & Klein, 2011). Complimentary, overall RTs are taken as a measure of monitoring abilities (Costa et al., 2009), and overall faster responses could be interpreted as better abilities to face demanding and changing contexts (bilingual advantage in monitoring or "BEPA" perspective in Hilchey & Klein, 2011).

As in the previous Chapters, the results of the tasks were analyzed following the traditional ANOVAs, as well as Bayesian Null Hypothesis testing comparisons (Rouder et al., 2009; Wetzels et al., 2011). Considering that the four tasks are going to be performed by the same set of participants, the correlation between the indices among the 4 tasks conducted here will be explored to see whether the inhibitory mechanisms applied to different tasks correspond to the same underlying construct or, on the contrary, the indices are completely independent from one another (Paap & Greenberg, 2013, Paap & Sawi, 2014).

II. *Experiment 8: The effects of bilingualism on the verbal Stroop task in young adults*

In this experiment, a group of young adults is tested to explore any potential differences due to bilingualism, using the widely studied Stroop test in its classical verbal version (Stroop, 1935). As it was explained in the previous chapters, this task allows for the assessment of both inhibitory skills (the *Stroop effect*) and monitoring abilities (global RTs). The inhibitory skills are measured by the difference between the congruent and incongruent condition, which shows how harmed the participant is when distracting irrelevant information (i.e., the incongruent condition), while the overall RTs represent how well participants deal with demanding and changing contexts, and are taken as an indicator of monitoring skills (see, for example, Costa et al., 2009). The *Stroop effect* has been used for these purposes in the literature by Bialystok and collaborators when they tested senior (around 68 years old) and young bilinguals (around 20 years old) in a Stroop task (Bialystok et al., 2008) . They found that bilinguals of both age groups displayed smaller *Stroop effect*, although the difference was more salient in the elderly group. Interestingly, the *congruency* effect (i.e., RTs in congruent trials as compared to RTs in control neutral trials) was larger and the *incongruity* effect (i.e., RTs in incongruent vs. control trials) was smaller in the bilingual sample than in monolingual sample (see Hernández et al., 2010, for a similar pattern), and again this was more evident in the old group.

If the previously reported instances of bilingual advantages in inhibiting and monitoring were not a consequence of uncontrolled external factors and small sample sizes (Paap & Greenberg, 2013), then I should be able to find the same significant pattern in the present study, where bilinguals and monolinguals were carefully matched in the potentially confounding factors discussed in the previous chapters (cf. Paap & Greenberg, 2013). This included general intelligence scores, socioeconomic status, immigration and ethnicity, so that the only relevant and distinctive factor between the two groups was their language profile.

1. Methods

a. Participants

180 young adults from Spain took part in this series of experiments. The 90 bilinguals (68 females, mean age 22.29 year, SD= 2.87) were recruited in Donostia-San Sebastian (in the Basque Country) and tested in the BCBL research center, in the same city. On average, they acquired Basque with 0.96 years of age (SD=1.27) and they reported to have a general proficiency of 8.41 over 10 (SD=1.88) in Basque. Their other language was Spanish, which was acquired with an average of 1.13 years (SD=1.72), with a mean punctuation of 8.58 (SD=1.91) based on self-reports. Thus, bilinguals were balanced in terms of proficiency (p>.33) and age of acquisition (p>.42). Apart from self-reported proficiency values (Clark, 1981; Heilenman, 1990; LeBlanc & Painchaud, 1985) and LexTale punctuations (see below), an interview conducted by a native speaker confirmed their mastery in Spanish and Basque. The 90 monolinguals (67 females, 21.84 years of age in average, SD=3.05) were recruited in the region of Murcia, in the south-east area of Spain, and tested at the University of Murcia. They reported to have acquired Spanish with a mean age of 0.68 (SD=.76), with the mean proficiency of 9.13 (SD=.84).

Participants from both groups were matched in a variety of factors that could potentially affect the experimental purposes, which have been shown to be of critical importance in the research of bilingual advantage (Morton & Harper, 2007; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015). Therefore, the 90 bilinguals and the 90 monolinguals were matched in age, IQ, socio-economic status (SES), educational level and knowledge of Spanish (see Table 15). An estimation of the IQ of each participant was based on their performance on an abridged version of the Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1990) that was administrated during the experimental session. As an indicator of the SES, total monthly income was considered and divided by the amount of household members, thus getting an approximate value of the incomes that each member of household receives monthly on average. Furthermore, regarding the educational level, the immense majority of the participants (88 bilinguals and 87 bilinguals) were highly educated, meaning that they already obtained a university degree (or higher) or were in the process of obtaining one. To control for their proficiency in Spanish, i.e. the language in which they all were tested, every participant completed the Spanish version of the LexTale task (Izura, Cuetos, & Brysbaert, 2014), thus providing with an objective indicator of their Spanish mastery. All these demographic and linguistic variables that could affect the outcomes of the study are thus controlled for (all *p*s>.1, see Table 15 for detailed information about the participants). All participants reported normal or corrected to normal vision and signed a consent form according to the principles established by the ethics committee of the BCBL.

Table 15.- Characteristics of the samples of monolingual and bilingual adults tested in Experiments 8, 9, 10 and 11. Mean values for each group are displayed with standard deviations between brackets. P values report independent sample t-tests comparisons' results.

	Monolinguals		**Bilinguals**		***p value***
Chronological age (years)	21.84	(3.05)	22.3	(2.87)	***0.31***
General IQ	22.76	(2.62)	23.4	(2.91)	***0.13***
SES (income in €/household members)	639.55	(498.97)	739.58	(297.36)	***0.1***
LexTale	92.28	(5.63)	93.4	(3.88)	***0.11***

b. Materials

The Spanish words for the colours red, blue and yellow ("rojo", "azul" and "amarillo") and three pairwise-matched (with a similar length, frequency and syllabic structure) non-colour words ("ropa", "avión" and "apellido", the Spanish words for clothes, plane and surname, respectively) were used. They were arranged to create the congruent (a colour word printed in the same colour that the word indicates; e.g., the word "azul" in blue), incongruent (a colour word printed in a different colour from what it is naming, e.g., the word "rojo" in blue) and neutral (non-colour words printed in any of the colours, e.g., the word "ropa" in red) conditions. 24 trials were used in each condition, and each colour was presented the same amount of times in each condition (8 times), and similarly each word was presented in the different colours the same amount of times. The order of the stimuli was randomized and there were no breaks in the experiment. All the strings were presented in uppercase Courier New font on a black background, while the colours were set in the RGB-scale values as follows: blue=0,0,255; red=255,0,0; yellow=255,255,0.

c. Procedure:

Bilingual participants were tested in the facilities of the BCBL in Donostia-San Sebastián, and monolingual participants were tested at the University of Murcia. In both locations, participants went through the experimental session in a room with equivalent settings and with the same equipment. The experiment was run using Experiment Builder (© SR Research), version 1.10.1385, and the CRT monitor was set to 60Hz in a resolution of 1280 x 1024. Sennheisser PC151 headsets were used to record participants' utterances.

Participants were instructed to name out loud the colour of the ink of the word on the screen as quickly and as accurately as possible. After a short training period, the experiment began. A fixation mark was presented for 250ms (a white cross centered in a black background), and then the target word appeared on the screen for 3000ms. Then it automatically moved to the next item until the experiment was finished.

2. Results:

For the analysis, audios were equalized to a 63dB amplitude using Praat© (Boersma & Weenink, 2015). Once all the files had same amplitude level, the voice onset was automatically detected by Praat as follows: each audio file was divided into "sound" and "silence" segments using the silence function from Praat. For a segment to be considered "sound" it had to have a minimum pitch of 100 Hz, to have exceeded a -25dB threshold and to have lasted at least 100ms. "Silence" segments had to last at least 200ms. The starting time point of the first sound segment was considered the onset of the speech and therefore, the reaction time of that response. The accuracy of the responses was checked manually, and the speech onset was manually adapted in the cases in which subjects corrected themselves (e.g., "roj…amarillo", Spanish for "re…yellow") and mistakes were removed.

a. Latencies

Before the analysis of the reaction times, outliers and errors were removed by deleting any response faster or slower than 2SD from the mean (4.84%). After this, a 3 x 2 general ANOVA was run with Condition (congruent, incongruent, neutral) as a within subject factor and Language (bilinguals, monolinguals) as a between subject factors (see Table 16). Condition was the only factor that resulted significant [$F(2, 356)=279.22$, $p<.01$], which showed that congruent condition was responded on average faster than neutral [$t(179)= 10.98$, $p<.01$]

and than incongruent condition [$t(179)= 21.32$, $p<.01$]. Neutral condition was also responded faster than incongruent condition [$t(179)= 13.80$, $p<.01$]. Crucially, no main effect of Language was observed [$F(1, 178)=1.53$, $p>.22$] and no interaction between it and Condition [$F(2, 356)=0.40$, $p>0.67$].

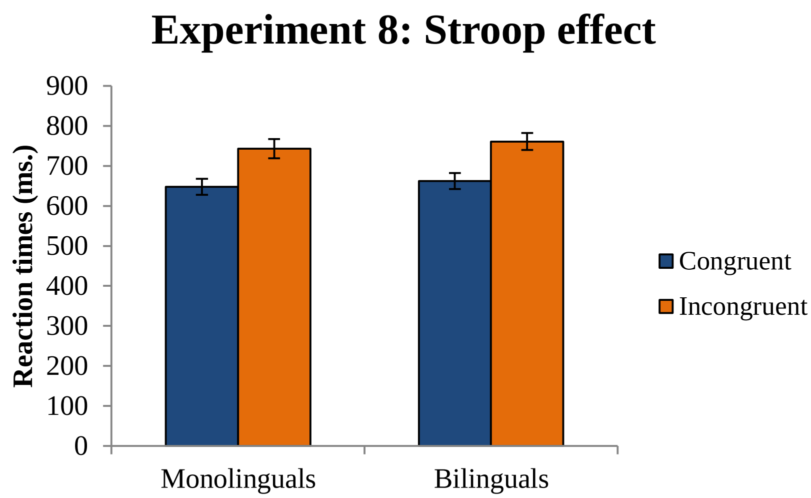

Figure 17. Reaction times for congruent and incongruent conditions in Experiment 8. Error bars represent confidence intervals of 95%

To further check the hypothesis of the bilingual advantage, a separate ANOVA was run for each of the indices. In the *Stroop effect* analysis (congruent vs. incongruent), a strong main effect of Condition was obtained [$F(1,178)=452.41$, $p<.01$], but the effect of Language was not significant [$F(1,178)=1.30$, $p>.29$] and, crucially, it did not interact with Condition ($F<1$). Importantly, the results of the ANOVA together with the Bayes Factor based on t-test comparison [$BF_{01}=5.62$] indicate that there are no significant differences between groups and that the null hypothesis is the most likely one to explain the data (see Table 16 for descriptive results, see also Fig. 17).

The *incongruity index* shows the exact same pattern, presenting a strong main effect of Condition [$F(1,178)=189.59$, $p<.01$] but negligible main effect of Language [$F(1,178)=1.77$, $p>.18$] and, importantly, no modulation of Condition by Language ($F<1$). Crucially, this null difference between groups was once again supported by the Bayesian t-test [$BF_{01}=5.79$].

Not surprisingly, *congruency index* once again replicates what was previously found: neutral items were responded to slower than the congruent condition [$F(1,178)=120.32$, $p<.01$] but this effect was not modulated by the knowledge of a second language [$F(1,178)=1.06$, $p>.30$], supported by the Bayesian t-test [$BFI_{01}=3.78$]. Similarly, no main effect of Language was found, indicating no overall RT differences [$F(1,178)=1.62$, $p>.21$].

Table 16.- Mean latencies for correct responses and error rates in all conditions for the verbal Stroop task ran in young adults in Experiment 8. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets. Indices for the effects are shown underneath the conditions

		Reaction times				Error rates			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	**Congruent**	648	(96)	662	(96)	0.05	(.44)	0.05	(.44)
	Incongruent	743	(116)	761	(104)	0.65	(1.97)	0.32	(1.43)
	Neutral	684	(86)	705	(98)	0.14	(.75)	0.14	(.75)
	Total	692	(94)	709	(95)	0.28	(.84)	0.17	(.72)
Effects	**Stroop**	95	(65)	99	(58)	0.6	(1.93)	0.28	(1.37)
	Congruency	35	(48)	43	(47)	0.09	(.88)	0.09	(.62)
	Incongruity	-60	(63)	-56	(50)	-0.51	(1.75)	-0.19	(1.24)

b. Accuracy

Error rate analysis showed a strong and significant Condition effect [F(2, 356)=10.24, p<.01], indicating that more errors were made in the items belonging to the incongruent condition than in the ones belonging to the congruent condition [t(179)= 3.52, p<.01] and to the neutral one [t(179)= 3.07, p<.01]. But, no effect of Language was found [F(1,178)= 0.87, p>.35], nor an interaction between Language and Condition [F(2, 356)= 1.67, p>0.19].

Exploring each index (see Table 16), the *Stroop effect* was strongly significant [F(1,178)=12.43, p<.01] but the difference between both conditions was not modulated by Language [F(1,178)=1.69, p>.2] and language groups did not differ in accuracy either [F(1,178)=1.35, p>.25]. The Bayes Factor comparison between the index across groups [BF_{01}=2,83] supported the null-differences hypothesis.

The *incongruity index* resulted significant [F(1,178)=9.45, p<.01] but Language factor did not modulate it [F(1,175)=2.06, p>.15] neither a main effect of Language was observed (F<1). Bayes factor comparison also tends to support the null hypothesis as the best fitting candidate [BF_{01}=2.38]

For the *congruency index,* Condition was not significant [F(1,175)=2.67, p>.1], and neither was it Language or the interaction between the two factors (all Fs<1).

As it can be seen from the results reported above, neither the latency nor the error rate indices are modulated by the linguistic profile of the participants, and this similarity between groups is also supported by the Bayesian Factor analysis.

III. Interim conclusions: Experiment 8

The Stroop task, one of the most extensively used tasks that tap into inhibitory control, was used to test bilingual and monolingual young adults. This task was previously used in seniors (cf. Chapter 2) and children (cf. Chapter 3), and the same results were replicated in the current Chapter with young adults. Namely, participants took longer to name the ink colour of words that would refer to different colours ("red" in green, for example) than to name a colour word's ink colour printed in that same ink ("green" in green) or to name the ink colour of a non-colour word . This was also the case for accuracy, where incongruent trials were more prone to error than neutral and congruent ones. The *Stroop, congruency* and *incongruity* indices expected from this task were strongly significant, but none of them varied depending on the linguistic profile of the participants.

Thus, previous experiments reporting a bilingual advantage on the Stroop task (Bialystok et al., 2008) were not replicated. Instead, when this task was tested in large samples of well-controlled bilingual and monolinguals (Paap, Johnson, & Sawi, 2015), the null hypothesis was undoubtedly preferred by the analyses that were conducted. These results go against the arguments supporting a bilingual advantage in inhibition (which should have produced a smaller *Stroop effect* in bilinguals) or monitoring (which should have been reflected in a main Language effect in RTs, indicating overall faster reaction times for bilinguals).

As it is going to be argued in the upcoming section, the verbal Stroop task can suffer from some flaws that makes it, on its own, un insufficient evidence for generalizing the absence (or presence) of a difference between bilinguals and monolinguals. Furthermore, given the low cross-task reliability of the bilingual advantage effects and its elusiveness in the sample of young adults, more tasks are required to explore the existence of the effect in this age range.

IV. Experiment 9: The effects of bilingualism on the numerical Stroop task in young adults

In Experiment 8, bilinguals and monolinguals did not differ in any of the indices that were taken as an indicator of inhibition or monitoring. As explained in the previous chapters, although it is one of the most extensively used tasks to tap into inhibitory abilities, the classic Stroop task comes with the inherent problem of being language based. This can be a problem when using it to compare monolinguals and bilinguals. That is to say, these two groups have been shown to differ in lexical access, with bilinguals consistently showing a poorer lexical access when spoken responses were required (Ivanova & Costa, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Thus, it can be difficult to isolate differences coming from inhibitory skills from the differences coming from basic linguistic performance variations. Therefore, in parallel to the experiments with the elderly and children, the numerical Stroop task was used in young adults, in order to study the impact of bilingualism on EF with a less linguistically charged task (Besner & Coltheart, 1979; Kaufmann et al., 2005; Santens & Verguts, 2011). This task requires that participants decide which of two presented digits is bigger in physical size, inhibiting the information about the actual numerical value represented by those numbers. Resembling the conditions present in the verbal version of the task, here congruent (e.g., a small digit 3 vs. a big digit 7), incongruent (e.g., a big digit 3 vs. a small digit 7), and neutral situations (e.g., a small digit 3 vs. a big digit 3) can be found. By comparing the results obtained in both verbal and numerical Stroop task, one could arguably be able to see whether the null effects obtained in the verbal task were just a product of the differences in the lexical access, or whether there are truly no differences between monolinguals and bilinguals in the EF tasks. In the former case, bilinguals and monolinguals should differ in the numerical Stroop task, and in the latter case, they should display no differences in performance.

The objective of this experiment is, firstly, to see whether the results from the verbal Stroop task can be replicated in a similar task that does not involve that much linguistic information, and secondly, to test whether the results obtained with adults resemble those obtained in the elderly and children. If bilingualism enhances inhibition ("BICA", Hilchey & Klein, 2011), a reduced *Stroop effect* in the bilingual group should be observed (Bialystok et al., 2008). If bilingualism, on the other hand, enhances monitoring ("BEPA", Hilchey & Klein, 2011), bilinguals should respond overall faster to all the conditions (Costa et al., 2009).

1. Methods

a. Participants

The participants were the same ones that took part in the Experiment 8 (see Table 15).

b. Materials

Stimuli consisted of six digits (2, 3, 4, 6, 7, and 8), arranged in pairs to form each trials (e.g., the digit 2 and the digit 6), one on the left side and another one on the right side of the screen. Depending on how the digits were paired, three conditions were created: 24 congruent trials (the number larger in value was also bigger in size, e.g., small digit 2 and big digit 6), 24 incongruent trials (the number larger in value was smaller in size, e.g., big digit 2 and small digit 6) and 24 neutral trials (two same numbers different in size, e.g., big digit 4 and small digit 4). In all the conditions "left" and "right" responses were equally distributed, and each digit was used in each condition the same amount of times.

c. Procedure

The general settings, installations, software and proceedings were the same as the ones followed in the Experiment 8. Manual responses were recorded using a response box with 7 buttons, and the participants used the leftmost and the rightmost button to give their responses.

Participants were instructed select the biggest number in size, ignoring their numerical value. After a short training period, the experiment began. A fixation mark was presented for 1000ms and then the target word appeared on the screen for 5000ms or until response was given. The order of the stimuli was randomized and there were no breaks in the experiment.

2. Results:

a. Latencies

After deleting the errors, the remaining latencies were cleaned for outliers by deleting any response faster or slower than 2SD from the mean (4.75%). A general 3 x 2

ANOVA was run including Condition as a within-subjects factor (congruent, incongruent, neutral) and Language as a between-subjects factor (bilinguals, monolinguals).

In the RT analysis, Condition was the only main effect that resulted significant [$F(2, 356)=202.38$, $p<.01$], and when this effect was analyzed in detail, congruent trials were responded to faster than both the incongruent ones [$t(179)= 16.37$, $p<.01$] or the neutral ones [$t(179)= 6.04$, $p<.01$], and that neutral items were also responded to faster than the incongruent ones [$t(179)= 13.80$, $p<.01$]. Crucially, no main effect of Language was found [$F(1, 178)=2.61$, $p>.11$], nor an interaction between it and Condition [$F(2, 356)=0.40$, $p>0.67$].

The *Stroop effect* (incongruent vs. congruent trials) resulted significant [$F(1,178)=268.63$, $p<.01$], but Language effect [$F(1,178)=2.95$, $p>.09$] and the interaction [$F(1,178)=1.44$, $p>.23$] indicated that the linguistic profile did not have any reliable impact (see Fig. 18), as confirmed by the Bayes Factor analyses [$BF_{01}=3.18$]. Not only there were no significant differences between groups, but the null hypothesis was three times more likely to explain the data than the alternative one (see Table 17 for descriptive results).

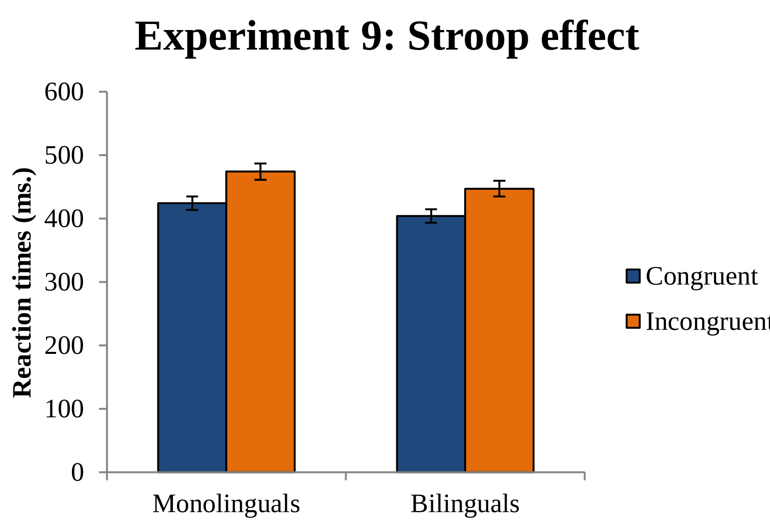

Figure 18. Reaction times for congruent and incongruent conditions in Experiment 9. Error bars represent confidence intervals of 95%

Similarly, the *incongruity index* (incongruent vs. neutral) analysis revealed that incongruent trials were responded to slower than the neutral ones [$F(1,178)=191.70$, $p<.01$], but Language effect was not significant [$F(1,178)=2.62$, $p>.11$], neither it was the interaction between them [$F(1,178)= 2.23$, $p>.14$]. The Bayes Factor analysis showed the tendency towards the null hypothesis [$BF_{01}=2.2$].

The *congruency index* (congruent vs. neutral) was strong [$F(1, 178)=36.25$, $p<.01$] but it was not modulated by Language ($F<1$), and Language itself was not significant either

[F(1,178)=2.19, p>.14]. Bayes Factor analysis undeniably favoured the null hypothesis as the best candidate [BF_{01}=5.89].

Table 17.- Mean latencies for correct responses and error rates in all conditions for the numerical Stroop task ran with young adults in Experiment 9. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets. Indices for the effects are shown underneath the conditions.

		Reaction times				Error rates			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	**Congruent**	424	(86)	404	(86)	0.37	(1.35)	0.37	(1.35)
	Incongruent	474	(103)	447	(100)	2.27	(4.29)	2.69	(4.43)
	Neutral	433	(95)	414	(91)	0.23	(1.15)	0.32	(1.56)
	Total	444	(93)	422	(90)	0.96	(1.75)	1.13	(1.68)
Effects	**Stroop**	51	(36)	44	(41)	1.9	(4.42)	2.31	(4.47)
	Congruency	10	(25)	11	(20)	-0.14	(1.70)	-0.05	(2.02)
	Incongruity	-41	(37)	-33	(35)	-2.04	(3.81)	-2.36	(4.72)

b. Accuracy

The error rate analysis also showed a significant Condition effect [F(2, 356)=38.79, p<.01], which was a reflection of incongruent trials producing more errors than both the congruent [t(179)= 6.26, p<.01] and the neutral [t(179)= 6.79, p<.01] ones, but no differences were found between congruent and neutral conditions(t<1). Importantly, no effect of Language or an interaction between Language and Condition (all Fs<1) were found.

The *Stroop effect* analysis revealed a main effect of Condition [F(1,178)=40.43, p<.01] but no other effects resulted significant (all Fs<1). Bayes Factor analysis [BF_{01}=4.80] indicated that both groups behaved similarly.

The same pattern was obtained for *incongruity effect*, showing a strong Condition effect [F(1,178)=47.32, p<.01] but no main effect of Language or interaction between the factors (Fs<1). Bayes Factor analysis, once again, indicated no differences in the behavior of the Language groups [BF_{01}=5.49].

In the *congruency index* analysis in error rates, none of the factors resulted significant (all Fs<1). However, a detailed Bayes Factor analysis indicated the same difference for both of the language groups [BF_{01}=5.88].

Thus, the null effects in the ANOVAs and the Bayes Factor values indicate that the indices obtained from both latency and error rate analyses were highly similar for both bilinguals and monolinguals, and no bilingual advantage was found whatsoever.

V. Interim conclusions: Experiment 9

In numerical Stroop task conducted in young adults, the pattern obtained in the previous verbal Stroop task was fully replicated. The results from the numerical Stroop task seem to suggest that the pattern obtained in the verbal Stroop task was not affected by differences in language processing between bilinguals and monolinguals (Ivanova & Costa, 2007; Gollan, Montoya, Fennema-Notestine, & Morris, 2005) as much as it could have been originally argued. On the contrary, these results go against previous results that indicate a bilingual advantage in inhibition or monitoring (see Bialystok, Craik, Klein, & Viswanathan, 2004; Martin-Rhee & Bialystok, 2004; see Hilchey & Klein, 2011, for review), and favours the opponent perspective that argues that any sign of an advantage fails to be replicated when the confounding factors are controlled for (see also Paap & Greenberg, 2013).

Importantly, the pattern of results obtained in verbal and numerical Stroop tasks completely resembles the results obtained in these two tasks with children and seniors, namely strong and reliable *Stroop effects*, displayed by a more effortful processing of the incongruent trials relative to neutral and congruent trials. This indicates that the task was sensitive enough to capture changes in responses related to the cognitive processes in question, in situations in which distracting information had to be inhibited (i.e., when the physical size and the numerical value of the numbers presented did not coincide) and in which the irrelevant information favoured the response.

It is worth keeping in mind that there are some studies that question whether the numerical Stroop task is completely free from any linguistic influence (Dehaene, 1992; Dehaene et al., 1993; Fias et al., 1996; but see Fias, 2001). In addition to this, the low cross-task replicability of the findings in young adults has recently been brought up (Paap & Greenberg, 2013). Therefore, the same hypotheses are going to be tested with two more tasks in the following sections, in order to gather more evidence that would strengthen (or refute) the findings observed in Experiments 8 and 9.

VI. *Experiment 10: The effects of bilingualism on the flanker task in young adults*

The results from the previous two experiments, the verbal and the numerical Stroop tasks, clearly indicated that bilinguals and monolinguals performed equivalently. The numerical task was used to try to capture inhibitory skills in a purer manner, i.e. the task is less language-charged than the verbal Stroop. However, it is worth mentioning that numerical Stroop's results might not be completely unaffected by verbal processing, given that some studies have indicated that number processing might be affected by semantic features (Dehaene, 1992; Dehaene et al., 1993; Fias et al., 1996; but see Fias, 2001 for results indicating asemantic written number word to phonetics translation). Bearing this in mind, I deem the inclusion of additional tasks convenient, and I implement the flanker task (Eriksen & Eriksen, 1974) with that purpose.

In the previous Chapter, the existence of the bilingual advantage in children was explored by means of the ANT task (Fan et al., 2002) as a language-free alternative to the Stroop tasks. This task explores the three attentional networks, namely *orienting, alerting* and *executive attention* (i.e., the *conflict effect*). As reported by Fan et al., (2002), "alerting is defined as achieving and maintaining an alert state; orienting is the selection of information from sensory input; and executive control is defined as resolving conflict among responses" (p.340) Although it is true that *alerting* and *orienting* networks have been explored in terms of their relation with bilingualism (see Costa et al., 2008; but see Costa et al., 2009; for differential findings in *alerting*; and see Colzato et al., 2008; but also Hernández et al., 2010; for differences in *orienting*), those networks should not be the ones that bilingualism would possibly affect. If bilingualism truly trains and enhances salient response inhibition, stemming from language competition, then the critical network to be explored is the one that the *conflict* index of the ANT task taps into. This network is explored by computing the difference between congruent and incongruent trials, and it would indicate how harmed the participants' responses are by the presence of conflicting information. In essence, the classic flanker task (Eriksen & Eriksen, 1974) is the *conflict* index of the ANT task. For this reason, and importantly in order to efficiently use participants' time and availability, only the flanker task (rather than the whole ANT task) was included in the experimental setting.

In a common version of this task, participants need to respond to a row of 5 arrows in the screen. They have to indicate whether the central arrow is pointing left or right, and that critical arrow can be flanked by another 2 arrows on each side, creating a congruent condition (all the arrows pointing towards the same direction, e.g., →→→→→) or an incongruent condition (flanking arrows pointing in the opposite direction; e.g., ←←→←←). Simple lines can also flank the central arrow, this way creating the neutral condition (e.g., - - → - -).

When the possible differences between bilinguals and monolinguals were explored using the flanker task (or the *conflict effect* in the ANT), an intriguing pattern has been found. As it was explained in the introduction to Experiment 7 (Chapter 3), Costa and colleagues (Costa, Hernández, & Sebastián-Gallés, 2008) found reduced *conflict effects* in the ANT when they tested young adults, but also faster overall reaction times. The latter finding was replicated later by the same team (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009), but only in highly demanding experimental contexts (with equal amount of congruent and incongruent trials), and no reduced *conflict effect* for bilinguals was found. Thus, it was argued that monitoring was the real enhanced component by bilingualism, and not inhibition (Bialystok et al., 2004; Green, 1998; Kroll, Bobb, Misra, & Guo, 2008; and see also Morales, Gómez-Ariza, & Bajo, 2013 for an explanation combining inhibitory and monitoring skills). Instead, the bilingual advantage in inhibition (see, for example, Pelham & Abrams, 2014; for a bilingual advantage in inhibition in young adults tested in the ANT) seems to be restricted to certain experimental conditions, often failing in its replication (see, among others, Kousaie & Phillips, 2012; Paap & Greenberg, 2013; and Prior & MacWhinney, 2010; for null results in the flanker task). These two perspectives are appropriately named as “BEPA” and “BICA” respectively by Hilchey and Klein (2011).

Hence, the present experiment aims at testing whether the lack of bilingual advantage in inhibition and monitoring found in the previous two experiments will be replicated in the present flanker task, or, on the contrary, whether bilinguals would be overall faster (thus indicating enhanced monitoring abilities) or would display less impact of the incongruent condition (i.e., enhanced inhibitory skills). Following the procedures reported in Chapters 2 and 3, as well as in the previous two experiments, the differences will be explored using both ANOVAs and Bayesian Null Hypothesis Testing.

1. Methods

a. Participants

The participants in this experiment were the same ones that took part in the Experiments 8 and 9 (see Table 15).

b. Materials

Rows of five arrows (←) were displayed on the center of the screen. For the congruent condition, the central arrow was flanked by four arrows pointing to the same direction (← ← ← ← ←). For the Incongruent condition, the central arrow was flanked by arrows pointing to the opposite direction (← ← → ← ←), and for the neutral condition, the arrow was not flanked by arrows (-- -- ← -- --). There were 16 items of each condition, 8 of them with the central arrow pointing to the left and the other half with the central arrow pointing to the right.

c. Procedure

The general proceedings, equipment and proceedings were the same as the ones used in Experiments 8 and 9.

Participants were asked to indicate with a button press the direction to which the central arrow was pointing. After a short training phase, the experiment started. First, a fixation point was displayed in the center of the screen for 1000ms in black, on a white background. Then, the row of arrows appeared on the screen for 5000ms or until the response was given. The order of the stimuli was randomized and there were no breaks in the experiment.

2. Results:

a. Latencies

After removing errors, latencies were trimmed for outliers by deleting any response that exceeded 2SD from the mean (5.15% of the data). After this, a 3 x 2 general ANOVA was run with Condition (congruent, incongruent, neutral) as a within subject factor and Language (bilinguals, monolinguals) as a between subject factor.

In the analysis of latencies, a strong main effect of Condition was observed [$F(2, 356)=196.15$, $p<.01$], and a more detailed analysis indicated that congruent items were responded to faster than the incongruent ones [$t(179)= 16.69$, $p<.01$], and also neutral items were responded to faster than both the incongruent [$t(179)= 15.98$, $p<.01$] and the congruent ones [$t(179)= 2.48$, $p<.02$] . There was no main effect of Language [$F(1, 178)=0.60$, $p>.44$] and no interaction between the two main effects [$F(2, 356)=0.01$, $p>0.99$]. To further check for any possible advantage, the indices were analyzed separately.

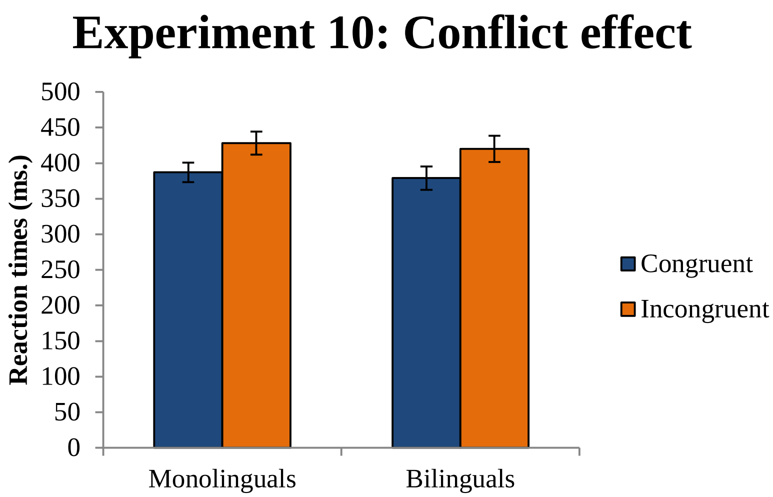

Figure 19. Reaction times for congruent and incongruent conditions in Experiment 10. Error bars represent confidence intervals of 95%

The *conflict index* analysis (i.e., incongruent vs congruent latencies) shows that Incongruent trials elicited slower responses than the congruent ones [$F(1,178)=279.92$, $p<.01$], but no other main effect or interaction was significant (Fs<1). Consequently, the Bayes t-test comparison indicated that the *Stroop effect* was equivalent between the two groups. [$BF_{01}=6.14$] (see Table 18 and Fig. 19 for descriptive results).

Table 18.- Mean latencies for correct responses and error rates in all conditions for the flanker task ran in young adults in Experiment 10. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets. Indices for the effects are shown underneath the conditions.

		Reaction times				Error rates			
		Monolinguals		Bilinguals		Monolinguals		Bilinguals	
Conditions	Congruent	387	(67)	379	(80)	0.49	(1.68)	0.49	(1.68)
	Incongruent	428	(78)	420	(89)	2.71	(4.40)	2.85	(4.97)
	Neutral	382	(64)	373	(68)	0.63	(1.89)	0.69	(2.19)
	Total	399	(67)	391	(77)	1.27	(1.97)	1.34	(2.03)
Effects	Stroop	41	(33)	41	(34)	2.22	(4.33)	2.36	(4.92)
	Congruency	6	(29)	6	(32)	0.14	(2.29)	0.21	(2.38)
	Incongruity	-46	(38)	-47	(41)	-2.08	(4.29)	-2.15	(5.48)

The *incongruity index* analysis perfectly mirrored the results obtained in the *Stroop effect*, showing strong Condition effect [$F(1,178)=253.96$, $p<.01$] but no other significant results (all Fs<1). Bayes factor comparison of the index between Language groups clearly favoured the null hypothesis as the best fit for the data [$BF_{01}=5.16$]

Similarly, *congruency index* analysis revealed that, even though Condition was significant [$F(1,178)=6.14$, $p<.02$], no other factor or interaction reached significance (Fs<1). Bayesian analysis once again favoured the null hypothesis as the best option to explain the index difference between language groups [$BF_{01}=4.70$]

b. Accuracy

The analysis of the error rates showed a similar pattern. A strong and significant Condition effect emerged [$F(2, 356)=34.39$, $p<.01$], stemming from incongruent trials producing more errors than the congruent trials [$t(179)= 6.66$, $p<.01$] and neutral trials [$t(179)= 5.79$, $p<.01$], but no differences were found when congruent and neutral conditions were compared [$t(179)= 1.00$, $p>.32$]. Importantly, no main effect of Language or an interaction between it and Condition was found (all Fs<1).

When the *conflict index* was analyzed, strong Condition effect was found [$F(1,178)=44.06$, $p<.01$], but no other main effect or modulation resulted significant (all Fs<1). Expectedly, the Bayes factor analysis [$BF_{01}=6.07$] supported the null-differences hypothesis.

The *incongruity index* also resulted significant in the error rate analysis [$F(1,178)=33.35$, $p<.01$], but it was not modulated by Language, and the main effect of Language did not result significant either (all Fs<1), which was confirmed by the Bayesian Factor analysis [$BF_{01}=6.16$].

The analysis of the error rates on the *congruency index* did not show any significant factor or interaction (all Fs<1), but still the index was compared across groups. Again, the null hypothesis was the best fit for the data [$BF_{01}=6.08$]

As in the previous tasks, the classic indices obtained from the latencies and error rates produced in the flanker task were not modulated by the language background of the participants, and the similarity between groups was strongly supported by the Bayesian Factor analysis.

VII. Interim conclusion: Experiment 10

The flanker task was included in this set of experiment to explore to what extent the results obtained from the verbal Stroop task (likely affected by lexical access differences between bilinguals and monolinguals) and the numerical Stroop task (much less affected by lexical access differences that bilinguals and monolinguals feature) are replicated in a task that relies on no semantic cues at all. This task was also included in the study with children in its extended version (ANT task, Chapter 2), and the same pattern was obtained. Given that the possible inhibitory abilities enhancement in bilinguals is basically captured by the *conflict index* of the ANT task, which is, in essence, the flanker task, it was decided to use only the flanker task in this experiment.

These data once again indicate that in a large and well-matched sample of bilinguals and monolinguals, the bilingual advantage is not captured in tasks that tap into different aspects of executive control (see, among others, Kousaie & Phillips, 2012; Paap & Greenberg, 2013; and Prior & MacWhinney, 2010; for null results in the flanker task) and suggests that previous significant findings might be due to uncontrolled external factors or small sample sizes (Pelham & Abrams; 2014).

It is important to notice that, in a large sample of well controlled monolingual and bilingual young adults, the results obtained from the flanker task replicate both the result obtained in the same task (within the ANT task) in children and, crucially, the indices obtained in the tasks tapping into the same underlying constructs in older adults. It must be noticed, also, that the Language main effects and interactions between Language and Condition were here even weaker than in the previous (which potentially could be more affected by extraneous factors such as lexical access differences) tasks. If previous research in this field indeed provides a solid evidence for bilingual advantage in inhibition, reduced *conflict indices* for bilinguals as compared to monolinguals should have been obtained (Costa et al., 2008). Similarly, if the monitoring advantage explanation is the one best reflecting the reality, a main effect of Language should have been found, indicating overall faster reaction times in the bilingual group (Costa et al., 2009). None of those results were obtained. Especially, it is worth mentioning that as several authors have claimed that the context should be demanding enough for the advantage in monitoring to appear (Bialystok, Craik, Klein, &

Viswanathan, 2004 and Martin-Rhee & Bialystok, 2008; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009), and that is how the tasks were implemented here, with a paradigm of equally present congruent, incongruent and neutral items, in a highly demanding context.

In order to improve the power of the generalizations and conclusions drawn for the present set of experiments, a final task is presented in the next section.

VIII. *Experiment 11: The effects of bilingualism on the Simon task in young adult bilinguals and monolinguals*

Finally, in Experiment 11, the set of participants tested in Experiments 8-10 was tested in the Simon task, one of the most extensively used tasks to study the bilingual advantage, in every age stratum. In the Simon task, each visual stimulus has an associated response (e.g., press "left" when you see a square and "right" when you see a circle), and congruent and incongruent trials are created by manipulating the position information that is irrelevant for the task itself (e.g., a circle presented in the left side represents an incongruent trial and a circle presented on the right represents a congruent trial). The difference in responses to congruent and incongruent condition has been taken as an index of how well participants deal with Incongruent and distracting information that needs to be inhibited. Crucially, it has been reported that the index is smaller for bilinguals than for monolinguals. Similarly to what has been reported for children (Bialystok, Martin, & Viswanathan, 2005) and the elderly (Bialystok, Craik, Klein, & Viswanathan, 2004), Bialystok (2006) found that the *conflict effect* (the difference between congruent and incongruent trials) was smaller for young adult bilinguals in the Simon task due their speeded responses in the incongruent condition (see also Bialystok, Craik, & Luk, 2008 and Bialystok & DePape, 2009). However, it is worth mentioning that the effect is not always strongly present: there are studies showing smaller conflict effects in the Simon task in young adults than in seniors (Bialystok, Craik, Klein, & Viswanathan, 2004). Furthermore, other studies report absence of behavioral differences between young adult monolinguals and bilinguals (Bialystok, Martin, & Viswanathan, 2005) or find neuroimaging, but not behavioral differences (Bialystok, Craik, Grady, Chau, Ishii, Gunji, & Pantev, 2003). Most of the studies showing significant effects of bilingualism are based on small sample sizes with uncontrolled external factors (Paap, Johnson, & Sawi, 2014) and, crucially, when things are better controlled for, the effects seemingly tend to vanish. Thus, Kousaie and Phillips (2012a) found no differences between young adult bilinguals and monolinguals in the Simon, flanker and the Stroop task and similarly, Paap & Greenberg (2013) found no bilingual advantage in flanker and Simon tasks.

Hence, this last experiment aims at testing whether the previously reported evidence in this thesis is trustworthy, or whether any difference between bilinguals and monolinguals in either inhibitory (which would be reflected in smaller *conflict effects*) or monitoring (which should create an overall faster responding pattern) abilities is finally captured.

1. Methods

a. Participants

The participants in this experiment were the same ones that took part in the Experiments 8, 9 and 10 (see Table 15).

b. Materials

A black square and a black circle were created, and by changing their position on a white background (centered in the vertical axis), three different conditions were created. The incongruent condition was created by presenting circles on the right side of the screen or squares on the left side of the screen, making participants respond to them using the button on the side opposite to the side of the screen in which the figure appeared (i.e., using the right button when detecting the square on the left, and using the left button when detecting the circle on the right). The congruent condition was created by presenting circles on the left and squares on the right side of screen, and making the participants respond using the response button on the corresponding side. Finally, the neutral condition was created by presenting the figures in the middle of the screen. There were 16 items of each condition, and half of each condition constituted squares, while the other half were circles.

c. Procedure

The procedure, equipment and peripherals were the same as the ones used in Experiments 8, 9 and 10.

Participants were presented with a black circle or a black square on the screen, and they were instructed to respond with the red button (on the left side of the response box) if they see a circle, or with the green button (on the right) if they see a square, irrespective of its position in the screen. After a short practice session, the experiment started with a black fixation point that was displayed in the center of the screen for 1000ms on a white background, and then the stimuli was presented on the screen for 5000ms, or until the response was given. When the response was given or the time limit was reached, the next item appeared on the screen. The order of the stimuli was randomized and there were no breaks in the experiment.

2. Results:

a. Latencies

After the errors were removed, the reaction time data were trimmed for outliers by deleting any response faster or slower than a 2SD from the mean (4.82% of the data). Then, a general 3 x 2 ANOVA was conducted including Condition (congruent, incongruent, neutral) as a within subject factor and Language (bilinguals, monolinguals) as a between subject factor.

Table 19.- Mean latencies for correct responses and error rates in all conditions for the Simon task ran in young adults in Experiment 11. Mean reaction times are showed in milliseconds with standard deviations between brackets. Error rates display percentages with the standard deviation between brackets. Indices for the effects are shown underneath the conditions.

		Reaction times				**Error rates**			
		Monolinguals		**Bilinguals**		**Monolinguals**		**Bilinguals**	
Conditions	**Congruent**	448	(124)	425	(112)	2.5	(4.37)	1.74	(3.63)
	Incongruent	478	(122)	457	(116)	4.38	(6.32)	4.03	(5.95)
	Neutral	460	(134)	433	(115)	2.36	(5.18)	2.01	(3.36)
	Total	462	(120)	438	(112)	3.08	(3.90)	2.59	(2.97)
Effects	**Stroop**	30	(63)	33	(39)	1.88	(7.34)	2.29	(6.45)
	Congruency	12	(73)	8	(36)	-0.14	(5.70)	0.28	(4.77)
	Incongruity	-18	(76)	-25	(41)	2.01	(5.84)	2.01	(5.99)

Latency analysis only revealed a significant main effect of Condition [$F(2, 356)=28.66$, $p<.01$]. Post-hoc comparisons showed that incongruent trials were responded to slower than both the congruent [$t(179)= 8.09$, $p<.01$] and the neutral ones [$t(179)= 4.71$, $p<.01$], and that congruent trials were responded to faster than neutral ones [$t(179)= 2.38$, $p<.02$]. However, neither main effect of Language [$F(1, 178)=1.9$ 1, $p>.17$], nor the interaction between them [$F(2, 356)=0.33$, $p>.72$] resulted significant (see Table 19 for descriptive results). To check for any possible difference between groups, the indices were compared across groups.

The *conflict index* was significant [$F(1,178)=65.01$, $p<.01$], but no main effect of Language [$F(1,178)=1.64$, $p>.20$] or an interaction was found [$F<1$], indicating that there are no significant differences in the *conflict index* across groups (see Fig. 20). Furthermore, when analyzing it with the Bayes t-test approach, the null hypothesis appears as the most likely one to explain the data [$BF_{01}=5.90$].

The analysis of the *incongruity index* showed the exact same pattern, with a significant Condition effect [$F(1,178)=22.01$, $p<.01$] but no effect of Language [$F(1,178)=1.89$, $p>.17$] nor interaction [$F<1$]. Bayes factor analysis also supported that the *incongruity index* was similar for both language groups [$BF_{01}=4.73$].

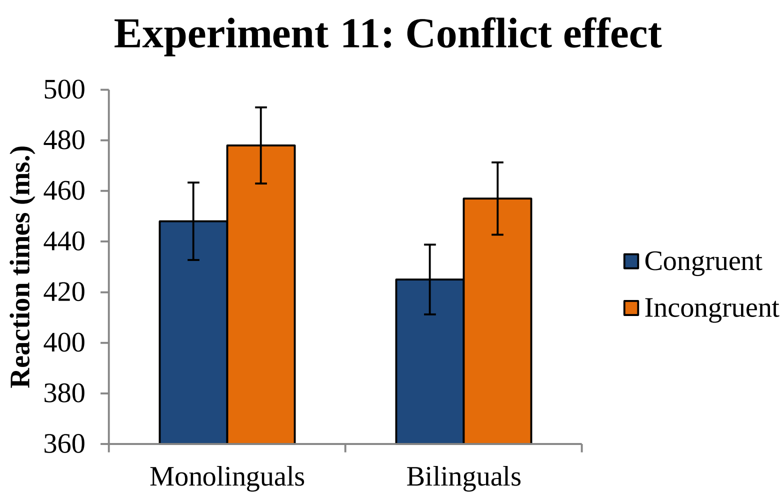

Figure 20. Reaction times for congruent and incongruent conditions in Experiment 1a. Error bars represent confidence intervals of 95%

Congruency index analysis resulted in a significant index [$F(1,178)=5.66$, $p<.02$] that did not interact with Language ($F<1$). Also, the effect of Language did not approach significance [$F(1,178)=2.09$, $p>.15$]. Bayes factor analysis supported the similarity of the indices across language groups [$BF_{01}=5.48$].

b. Accuracy

A similar pattern emerged for the error rates. Condition resulted significant [$F(2, 356)=13.7$, $p<.01$], and paired comparisons revealed that it was due to the incongruent trials triggering more errors than both congruent [$t(179)= 4.05$, $p<.01$] and neutral ones [$t(179)= 4.58$, $p<.01$], with no difference between the last two ($t<1$). No effect of Language or an interaction between it and Condition were found (all *F*s<1).

Crucially, the *conflict index* appeared significant [$F(1,178)=16.35$, $p<.01$], but it did not interact with Language and also, Language did not approach significance (all *F*s<1). Bayes factor analysis [$BF_{01}=5.74$] indicated that the null hypothesis was almost 6 times more likely to explain the data, thus suggesting that the indices were highly similar across Language groups.

The ANOVA analyzing the *incongruity index* revealed only a main effect of Condition [$F(1,178)=20.88$, $p<.01$], but neither Language nor the interaction between it and Condition resulted significant (all Fs<1). In line with this, Bayes factor analysis indicated that the *incongruity index* was highly similar across the language groups [$BF_{01}=5.79$].

Finally, the ANOVA on the *congruency index* revealed no main effect of Condition, Language, neither an interaction between them (all Fs<1.3, all ps>.26).

As in the previous tasks, not a single index showed a modulation based on the language profile of the participants, and the Bayes Factor analysis always supported the null hypothesis as the best fit for the data.

IX. Interim conclusion: Experiment 11

The results of the Simon task fully replicated the other 3 experiments that were conducted in young adults. Even though the *conflict effect* in the Simon task (the difference in RTs or accuracy between the congruent and incongruent condition) has been reported to be smaller for bilinguals than for monolinguals in children (Bialystok, Martin, & Viswanathan, 2005) seniors (Bialystok, Craik, Klein, & Viswanathan, 2004, in elderly) and young adults (Bialystok, 2006; Bialystok, Craik & Luk, 2008 and Bialystok & DePape, 2009), I hereby demonstrated that those differences tend to vanish (Kousaie & Phillips, 2012a) when samples are big enough and confounding demographic variables are controlled for (Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015a). Similarly, even though some researchers have argued for an advantage in monitoring abilities that would be reflected in faster overall reaction times (Costa et al., 2009), the bilinguals and monolinguals tested here showed equivalent overall reaction times, indicating equivalent monitoring abilities.

X. General discussion: Bilingual and monolingual young adults

The four experiments in this Chapter aimed at exploring the potential effects of bilingualism on executive functions in young adults. Carefully matched large cohorts of bilinguals and monolinguals of the same country were tested in the 4 mayor tasks that have been classically used to measure executive functioning skills. Crucially, all the relevant demographic factors (age, IQ, SES, educational level or immigrant status) were controlled for, by testing young adults that share historical, cultural and social background but differ in linguistic profile, in order to explore the reliability and replicability of the previous findings on children and adults.

Being aware of the low cross-tasks replicability and the instability of the bilingual advantages in young adult samples (see Paap & Greenberg, 2013), the main classic psychological tasks that tap into inhibitory abilities were used in the present chapter to see whether any previously reported advantage was replicated in the large cohort of carefully matched participants reported here. The same set of 180 participants went through flanker (Eriksen & Eriksen, 1974), Simon (Simon & Rudell, 1967) and two versions of the Stroop (Stroop, 1935) tasks.

As Hilchey and Klein (2011) comment, some researches have argued that bilingualism enhances the performance in situations where irrelevant information needs to be inhibited (the advantage in inhibition), while others defend that it is monitoring what gets enhanced by bilingualism. Accounting for the first theory, previous research has reported reduced *conflict indices* in Stroop (Bialystok et al., 2008), flanker (Costa et al., 2008) and Simon tasks (Bialystok, 2006; Bialystok, Craik, & Luk, 2008; Bialystok & DePape, 2009), while the defenders of the second argument have shown that bilinguals showed an overall advantage if the experimental setting is demanding enough (Bialystok, Craik, Klein, & Viswanathan, 2004 and Martin-Rhee & Bialystok, 2008; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009). However, the new line of research in this field argues that uncontrolled external factors such as intelligence, SES or immigrant status might be behind these differences (Paap, Johnson, & Sawi, 2015, see, for example, participants' data in Bialystok and Martin, 2004, Bialystok and Shapero, 2005; Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; and Bialystok, Craik, & Luk, 2008). Consequently, the differences in the very same tasks disappear when those factors are controlled for in large samples (Paap & Greenberg, 2013). Among others, Paap & Sawi (2014) found no evidence for bilingual advantage using the Simon and the ANT tasks, as well as antisaccade and colour-shape Stroop task. Kousaie and Phillips (2012a) failed to find bilingualism-related benefits in the Simon, Flanker and the Stroop task. Hence, the results I present in the present chapter add to these studies that clearly suggest that no differences due to bilingualism emerge in when large groups of monolinguals and bilinguals are matched in such a way that linguistic profile is the only difference between them.

It is worth noting that every task produced the expected classic patterns of results, with strong and constant *conflict effects* for all the participants. Each condition (incongruent, congruent and neutral) behaved as expected and in accordance with the previous literature. Nevertheless, there was no trace of neither better inhibitory skills, which would have been represented by a reduction of the *conflict effect* and indicated by an interaction between Language and Condition, nor better monitoring skills, which would have been evidenced by a main effect of Language in the ANOVA, reflecting overall faster reaction times. The lack of differences in *conflict effect* between language groups indicates that the performance decline produced by the incongruent trials as compared to the congruent ones did not vary between monolinguals and bilinguals, showing that both samples feature equivalent inhibitory skills. Complimentarily, bilinguals were not overall faster, which indicates that bilinguals' and monolinguals' monitoring abilities do not differ. Furthermore, the much more restrictive Bayesian factor analysis clearly shows that the null hypothesis of no differences between monolinguals and bilinguals is the best explanation for the results obtained, reinforcing the standpoint the bilinguals and monolinguals perform similarly in tasks that tap into EF skills (see Table 20 for a summary).

When four tasks that are tapping into mechanisms sensitive to inhibitory and monitoring abilities show the same patterns in a sample of well-controlled adults, the conclusions that can be derived are strong. However, even if it can be concluded that young adult bilinguals and monolinguals do not differ from each other in inhibitory or monitoring abilities as measured by these 4 tasks, still the question of whether those four tasks tap into the same question remains.

Similarly to the analysis performed in elderly and children, the cross-task coherence was tested in a correlation analysis, by comparing all the *conflict/Stroop indices* obtained in all the tasks (Paap & Greenberg, 2013; Paap & Sawi, 2014). Analysis indicated that the *Stroop/conflict effects* across tasks showed negligible correlation strength (all rs between -.06 and .10). *Congruency effect* showed some correlations worth considering (i.e., flanker task and numerical Stroop showed a correlation of r= .21), while all the others reflected mild cross-task reliability (all rs between -.05 and .15). Finally, *incongruity effects* showed the same pattern, with some considerable but negative correlation (flanker task and Simon task showed a correlation of r= -.20,) while all the other analyses indicated that the relation between the indices across tasks was not strong (all rs between -.11 and .17).

Table 20.- Update and summary of the results obtained in the present chapter. A cross indicates non-significant effects of language in the measures obtained. Correlations are reported only if significant.

	Tasks used	Inhibition	Monitoring	Between-task correlation
Elderly	*Verbal Stroop*	✕	✕	Incongruity in Exp. 3-4, r= -.52 Rest, n.s.
	Numerical Stroop	✕	✕	
Children	*Verbal Stroop*	✕	✕	N.s.
	Numerical Stroop	✕	✕	
	ANT	✕	✕	
Young Adults	*Verbal Stroop*	✕	✕	Congruency effect between
	Numerical Stroop	✕	✕	Flanker and numerical Stroop, r= .21
	Flanker	✕	✕	Incongruity effect between flanker and Simon, r= -.20
	Simon	✕	✕	Rest, n.s.

Seemingly, these results indicate that, although the tasks implemented here are all based on the same principles of presenting congruent and incongruent data, and can be interpreted as a measure of inhibitory abilities, their generalization to domain general abilities – or other domains is still uncertain. If the tasks were all tapping into the same underlying mechanisms, the correlation analysis should have shown strong between-task correlations.

Chapter 5: General discussion

In the current thesis I presented various experiments that aimed at exploring the replicability of the so-called bilingual advantage by testing large samples of children, young adults and seniors (see Fig. 21). With that purpose, a large variety of tasks that have been used previously in the literature to report significant differences between bilinguals and monolinguals of different ages were run. In the present collection of experiments no evidence of a bilingual advantage was found whatsoever.

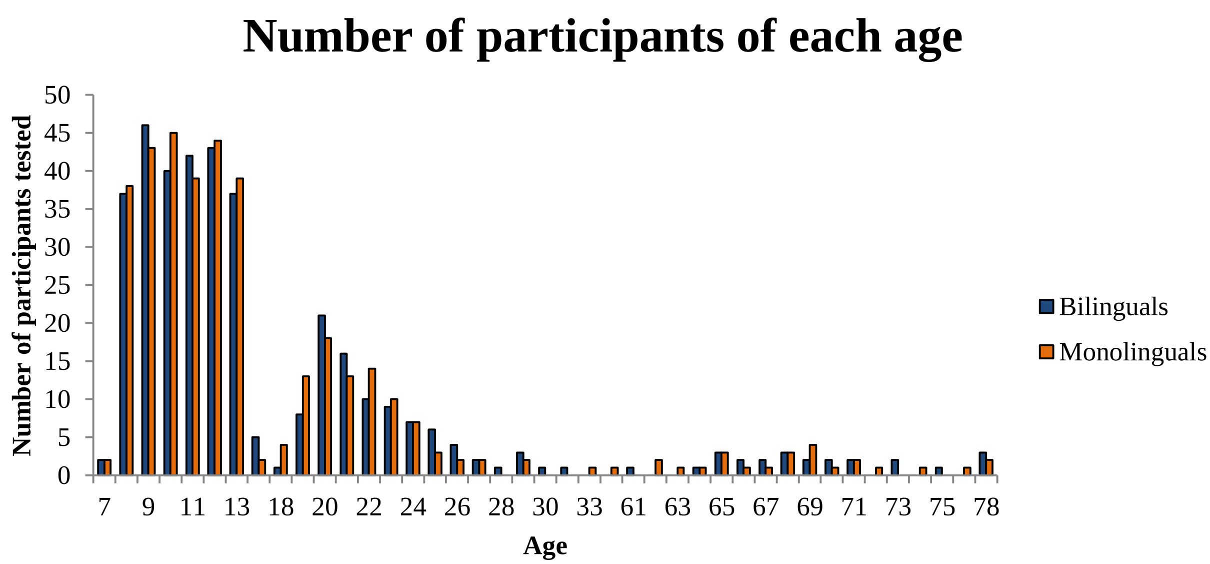

Figure 21. Absolute number of participants of each age tested in the present thesis.

When the bilingual advantage is defined, two main perspectives are usually taken (see, among others, Hilchey & Klein, 2011): whilst one of them argues for an advantage on inhibitory control mechanisms, the second postulates that the advantage comes from enhanced general executive processing and is usually equated with better monitoring abilities. Both of them stem from the same repeatedly reported fact that the two languages that a bilingual speaks are always active (Kroll, Dussias, Bogulski, & Kroff, 2012; van Hell & Djikstra, 2002) and are likely to affect each other (MacLeod, 1991; Smith, 1997; Ehri & Ryan, 1980; Miller & Kroll, 2002; Marian & Spivey, 2003; Hartsuiker, Pickering, & Veltkamp, 2004; Flege & Port, 1981; Hermans, Bongaerts, de Bot & Schreuder, 1998; Hernandez & Reyes, 2002; Costa, Caramazza, & Sebastián-Galles, 2000; Hoshino & Kroll, 2008). Consequently, bilinguals would need to constantly monitor their language system (with competing languages) to be able to choose what is required for the context while also monitoring the linguistic needs of the environment, eventually making them better at monitoring in general terms. That would account for the general executive functions enhancement in monitoring. If monitoring is important, so it is inhibition: once the contextual needs are identified, the non-relevant target must be suppressed to avoid any cross linguistic interference (Green, 1988). That would arguably enhance their domain-general inhibitory abilities.

These two mechanisms respond to some needs that are present in bilinguals (i.e., language control) but not in monolinguals, and that is why they might eventually enhance the above mentioned concrete aspects of executive functions. As argued by the bilingual advantage hypothesis, those enhancements would not only be restricted to language domain, but they would extend their effect to domain-general situations. Thus, when bilinguals face domain-general situations in which changing demands have to be monitored (for example, in a situation with intermixed congruent and incongruent stimuli), an enhanced general executive processing (equated to monitoring) is sometimes reported as faster reaction times in tasks with fluctuations in response relevant stimuli. Similarly, in domain general tasks that contain stimuli involving confounding and salient information, i.e. information that has to be inhibited because it favours the opposite response from the one intended (incongruent trials), better inhibitory abilities are reported by some studies. This indicates that bilingual participants are less harmed by incongruent stimuli as compared to the congruent ones (stimuli which contain irrelevant information that also favours the intended response).

In the present thesis, when the same hypotheses were explored in the most extensively used tasks, none of these hypotheses were confirmed in the participant samples at scrutiny here. Every statistical test run to find evidence of differences eventually favoured the null hypothesis: bilinguals and monolinguals, on the tasks they were tested on, showed the same monitoring and inhibitory abilities.

I. Review of the results

a. Seniors

Exploring each demographical group's outcome separately, I began by presenting data from elderly participants in the second chapter, where I analyzed the effects derived from lifelong bilingualism in domain-general cognitive.

Previous evidence was mixed, given that some research shows that bilingual seniors outperform monolinguals (Bialystok, Craik, & Luk, 2008; Bialystok, Craik, Klein, & Viswanathan, 2004; Gold, Kim, Johnson, Kryscio, & Smith, 2013), but recently other researchers have suggested that those pieces of evidence are not entirely reliable and replicable (Kirk, Fiala, Scott-Brown, & Kempe, 2014; Kousaie & Phillips, 2012; de Bruin, Bak, & Della Sala, 2015). In order to delve into this issue, the hypothesis of the bilingual advantage was tested in two ways. Firstly I used the verbal and the numerical Stroop tasks in a direct comparison between bilinguals and monolinguals (Experiments 1 and 2, respectively). Secondly, I presented data from two experiments where I tested bilinguals of different proficiencies (from almost monolingual to fully balanced bilinguals) in the same tasks, to look at any possible modulation of any index by the level of bilingualism (Experiments 3 and 4).

The reason why I first looked at the possible effects of bilingualism on this sample is because, if bilingualism counts as a training that would eventually improve inhibitory control and/or monitoring skills, its effect should be stronger in a population that has been "under training" their whole life (i.e., lifelong elderly bilinguals). Furthermore, these effects should be more easily captured in a population that is not at the maximum of their cognitive abilities and therefore any boosting or additional benefits would be more salient. It seems coherent to assume that the participants who took part in the studies were presumably undergoing 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 and 2 (Median = 29.5), and above 24 in Experiment 3 and 4 (Median = 29)). Attending to the voices that have recently been raised against the bilingual advantage, arguing that it is generally a consequence of unmatched samples at test (Paap, Johnson, & Sawi, 2015a), the samples were painstakingly controlled: every single participant of the sample had been part of the same society from birth, and bilinguals had used both of their languages over the vast majority of their lives on a daily basis. None of them was an immigrant, and they all had always lived in the same city. Importantly, bilinguals and monolinguals did not differ on any relevant demographic factors or in their linguistic skills in Spanish (which was the shared language between bilinguals and monolinguals and the language in which they were tested, see methods section, Chapter 2).

In Experiments 1 and 2, where bilinguals and matched monolinguals were compared in the verbal and numerical Stroop tasks, the results unambiguously show a complete absence of differences between lifelong bilingual seniors and their monolingual peers either in monitoring abilities (which means that they do not differ in overall reaction times) or in inhibitory skills (indicated by the fact that they are equally affected by the *Stroop effect*s). Importantly, when the same analyses were run including only those participants that arguably had an educational level for which the effects of bilingualism on cognition are more salient in the elderly (see Gollan et al., 2011), the same results were replicated, demonstrating that the bilingual advantage is not circumscribed to other factors.

In Experiments 3 and 4, where a large group of bilinguals that differed in their L2 proficiency was tested on the same tasks, none of the indices obtained correlated with individuals' L2 and no modulation of bilingualism was observed on any index of inhibition or monitoring. Furthermore, when the analyses were repeated including covariates that could affect the final outcome, such as IQ (Adelman et al., 2002; Arffa, 2007) or Education (Houx, Jolles, & Vreeling, 1993; Moering, Schinka, Mortimer, & Graves, 2004; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006), and despite the normal and expected modulation of the effects based on those covariates, knowledge of the second language did not modulate any index.

The general picture in seniors, as replicated in the four experiments, is straightforward: despite the fact that the indices obtained were comparable in size to previous findings in the literature for both bilinguals and monolinguals, none of the Stroop indices, none of the overall RT measures and, in a more detailed analysis, even when the congruency and *incongruity effects* were analyzed separately (see Bialystok et al., 2008; see also Barch et al., 1999; Carter, Mintun, & Cohen, 1995), no differences were observed between the two linguistic groups (note that Bayes factor analysis favoured the null hypothesis) nor did L2 proficiency modulate them.

The results from Chapter 2 conclude that lifelong bilingualism does not provide inhibitory or general executive processing advantages to elderly bilinguals as compared to well-matched elderly monolinguals. Furthermore, it also discredits the idea that bilingualism, as a continuum, can modulate the outcomes of tasks tapping into different aspects of executive functions. Based on these results, the next chapters focus only on comparisons between bilinguals and monolinguals (as an "all or nothing" factor) on the same tasks that the elderly were measured on and, additionally, other domain general executive functioning tasks.

b. Children

In Chapter 3, I present data resulting from testing the bilingual advantage in the opposite tail of the age distribution by comparing bilingual and monolingual children in the two versions of the Stroop task that were used with elderly. Additionally, the ANT task was included in the experimental set, to have a measure of executive functions with no linguistic influence.

The classic Stroop task (Experiment 5) requires the direct selection of a lexical item (i.e., the name of the colour to be verbally produced). In principle, it could be assumed that the numerical Stroop task (Experiment 6) 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 (i.e., bilingual children of this study 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. Concretely, 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 debatable.

Considering the fact that that both Stroop tasks (although to different degrees) could arguably be affected by lexical access, I deemed necessary the inclusion of a task that taps into executive functions with no language involvement. The third experiment of this chapter (the ANT, Experiment 7) was added with the purpose of disentangling that specific issue by comparing bilingual and monolingual children in a much less linguistically charged task. Furthermore, it would allow understanding the differences between groups (if any) in the four attentional networks that the ANT is tapping to. On the one hand, if lexical access differences masked the differences that executive functions could have caused in both of the Stroop tasks, these should be captured with the ANT task. If the similar performance of both language groups was reliable, null differences should be replicated in the ANT task also, evidencing no hidden beneficial effects in the previous Stroop tasks. Also, it must be borne in mind that there is a debate between researchers defending a better performance of bilinguals in the ANT task (e.g., Kapa & Colombo, 2013; Pelham & Abrams, 2014) and those suggesting that it is restricted to certain conditions and designs (e.g., Costa et al., 2009), so further investigation to see whether a large sample of bilingual children would exhibit better performance in this task than a group of carefully matched monolingual children was needed.

In the three tasks used to test the potential benefits of bilingualism in executive control in children, i.e. two Stroop tasks and the ANT, the evidence favouring a clear-cut difference between monolinguals and bilinguals in inhibitory skills was absent. Bilinguals did not outperform monolinguals in any of the tasks and any of the indices, and therefore no bilingual advantage was found whatsoever in either inhibitory or monitoring indicators.

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. The data show unequivocal *incongruity* and *congruency* effects across the whole range of ages under test. In both the Stroop tasks (Experiments 5 and 6), which were analyzed similarly to those results of elderly, 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*). Similarly to what happened with bilingual seniors, both the classic and numerical versions of the Stroop task (Experiments 5 and 6, respectively) showed a significant generalized *Stroop effect* (incongruent vs. congruent trials), but the size of the effects was similar for both bilinguals and monolinguals. Crucially, Bayes factor analysis clearly favoured the null hypothesis as the best explanation for the results obtained.

When the analyses were based on the error rates, the pattern observed closely resembles that obtained in the reaction time analyses. The results, again, offer a picture that is inconsistent with the proposal of enhanced cognitive control in bilingual children, either in general speed enhancement or in inhibitory abilities. Only a small difference was found in the classic verbal Stroop task (Experiment 5), when the congruency effect was compared and bilingual children showed a slightly larger effect than their monolingual peers (see also Bialystok et al., 2004). This difference does not support the perspective of the bilingual advantage situated in enhanced inhibitory abilities (there is nothing to inhibit in the congruent and neutral condition, nothing that interferes with the intended response). It could be accommodated by the perspective that posits the bilingual advantage in a general executive processing enhancement, producing better monitoring abilities rather than in inhibitory mechanisms (e.g., Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; Costa et al., 2009). However, this result should be taken with extreme caution because it was not replicated in the subsequent numerical Stroop task (Experiment 6) and, importantly, in any other task featured in the present thesis, with no difference in any congruency index obtained in any task in any population. Importantly, if a general monitoring enhancement were to be produced by bilingualism, it should have been reflected in overall faster reaction times and not only in error rate modulation of the *congruency index*. If we consider the large sample sizes and the reduced error percentage that they showed, I deem the degree of generalization of a potential bilingual advantage based on this error effect very limited.

In a nutshell, the interactions between the magnitudes of these indices (*Stroop* and *incongruity effects*) classically associated with inhibitory control and the linguistic profiles of the participants (monolinguals vs. bilinguals) did not unequivocally support a bilingual advantage in inhibitory skills in the two Stroop tasks. General group reaction times, which would have indicated different monitoring abilities, did not differ either. Furthermore, when the indices were compared across groups using Bayesian t-test, it clearly favoured the null hypothesis.

The results obtained from Experiment 7 clearly indicated that the so-called bilingual advantage could not be replicated with the ANT when a sufficiently large and well-matched group of bilingual and monolingual children were tested. I argue that if the so-called bilingual advantage were a consequence of bilinguals' enhanced inhibitory skills, a reduced *conflict effect* should have been found for the bilingual group (i.e., smaller differences between incongruent and congruent trials for bilinguals than for monolinguals). This was not the case, and participants performed in a highly similar fashion in these two conditions regardless of their linguistic profile. On the other hand, if the previously reported bilingual advantage was the result of bilinguals' enhanced monitoring skills, one would have expected an overall difference between groups in the RTs and/or in the error rates (e.g., Costa et al., 2009; see also Wu & Thierry, 2013), but again no supporting data for this claim was found. Once again, the Bayes factor analyses support the perspective of both linguistic groups behaving similarly.

It is worth mentioning that the task used was sensitive enough to all the indices tested and the expected developmental patterns. Preceding evidence from the monolingual domain was replicated, with both linguistic groups showing a *conflict* effect (longer reaction times and error rates for incongruent than for congruent trials), an *alerting* effect (a better performance in double cue trials as compared to no cue trials), an *orienting* effect (central cue trials responded slower and less accurately than valid cue trials) and a *validity* effect (longer RTs and higher error rates for invalid cues than for valid cues). This replicates the indices previously observed in the same task (e.g., Fan, McCandliss, Fossella, Flombaum, & Posner, 2005; Fan & Posner, 2004; Ishigami & Klein, 2010; Mackie, Van Dam, & Fan, 2013; Wang & Fan, 2007; Yin et al., 2012 among many others), so it is unlikely that the group differences were masked due to a lack of statistical power. Results also replicated previous findings from a developmental point of view (Rueda et al., 2004): the *conflict* effect, the main index of interest, diminished a lot as a function of age, *validity* and *orienting* showed modest changes and *alerting* suffered no modulation by age.

In a nutshell, all the tasks used to analyze the bilingual advantage in children yielded null results, and all the analysis converge in the null hypothesis of no differences in inhibition or monitoring to explain the data I presented.

c. Young adults

The objective of Chapter 4 was to explore the potential effects of bilingualism on executive functions in young adults, by conducting large scale tests with all the major tasks used in this regard in the literature. Therefore, the two versions of the Stroop task were included together with the Simon and the Flanker task. The decision of including the Flanker task and not the ANT was based on efficiency: there was no modulation of any index of the ANT by bilingualism on children, but the crucial one that, according to the bilingual advantage in inhibition, should be modulated by a knowledge and use of a second language would be the *conflict effect*. In essence, the flanker task is sufficient to observe the *conflict effect*, because it presents participants with congruent and incongruent (the ones to which inhibition needs to be applied) trials. Similarly, the Simon task was added to account for the low cross-task *validity* that has been often reported (Paap & Greenberg, 2013) and have a more complete picture of how bilingualism can affect to these broadly used tasks. Hence, large carefully matched cohorts of bilinguals and monolinguals of the same country went through all the tasks mentioned above, which have been classically used to measure executive functioning while relevant demographic factors (age, IQ, SES, educational level or immigrant status) were controlled for.

At first, previous evidence is seemingly consistent: advantages have been found in Simon (Bialystok; 2006 ; Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok & DePape, 2009), Simon and Stroop (Bialystok, Craik, & Luk, 2008) and flanker tasks (Abutalebi et al., 2012, Costa et al., 2008, Marzecová, Asanowicz, Krivá, & Wodniecka, 2013 Kapa & Colombo, 2013; Pelham & Abrams, 2014). However, upon careful review it is clear that these studies were based on small and uncontrolled samples (see for example Bialystok & Martin, 2004, Bialystok & Shapero, 2005; Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; Bialystok, Craik, & Luk, 2008), which led to some replication failures when those critical factors were controlled for (Morton & Harper, 2007; Hilchey & Klein, 2011; Kousaie & Phillips, 2012a; Paap & Greenberg, 2013, Paap & Sawi, 2014, Paap, Johnson, & Sawi, 2015a).

As Paap and his colleagues (2013) argue, the bilingual advantage should be present in various tasks that tap into the same executive functions (if tested in the same population) to consider it real, and that is why large cohorts of participants were tested in the four tasks mentioned above while controlling for the reported influencing demographical factors (Paap, Johnson, & Sawi, 2015a). The results are clear again: The *Stroop or conflict effects* should have been reduced in bilinguals as compared to monolinguals if bilingualism enhances inhibitory skills (Bialystok, 2011), whereas bilinguals should have shown overall faster reaction times if the key factor is an improvement in monitoring capacities (Costa et al., 2009). While previous research shows significant results using these very same tasks (e.g., Bialystok et al., 2008, Bialystok, 2006; Costa, Hernández, & Sebastián-Gallés, 2008), in the studies I report they yielded null effects. The tasks worked as expected, showing the classic patterns of strong and constant *conflict effects* mainly due to *incongruity effects*, and all the conditions behaved in accordance to the previous literature. However, none of these things was modulated by the knowledge of a second language. Once again, Bayesian factor analysis clearly indicates that the null hypothesis explains the data much better than any alternative hypothesis that claims for differences. Complimentarily, the monitoring-based advantage hypothesis is discarded due to similar global RTs that both language groups showed.

II. No bilingual advantage: Summary and proposals

The data I show in this thesis could be read as an indicator of the null developmental impact that bilingualism has on executive functions. It indicates that neither in children, nor in adults nor in the elderly, did bilingualism improve inhibition or monitoring.

a. Ceiling effect due to age variations on cognitive abilities

One might wonder why I do not find a strong and stable bilingual advantage in all the employed tasks. The results found here add to the growing body of evidence supporting a perspective that has been gaining strength in the last several years. This view suggests 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 observe bilingual advantages. The absence of evidence favouring the bilingual advantage in young adults (Bialystok, Martin, & Viswanathan, 2005, Chapter 4 of the present thesis) has been argued to be a consequence of the ceiling effects that cognitive skills might possess at those ages; and therefore it might be captured more easily in children or the elderly. However, this has not been the case when contrasting a large sample of monolingual and bilingual children (c.f. Chapter 3, see also Gathercole et al., 2014). 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 the benefits provided by bilingualism 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 (who have been exposed to bilingualism their entire life) whose cognitive skills are presumably declining (avoiding thus the ceiling effect). Nonetheless, here in Chapter 2 I demonstrated that in the elderly bilingual sample the bilingual advantage is equally absent (see also Kirk et al., 2014; Kousaie & Phillips, 2012; de Bruin, Bak, & Della Sala, 2015). Closing this circle, when the same hypothesis is tested in young adults (Chapter 4), bilinguals and monolinguals showed equivalent inhibitory and monitoring abilities.

In a situation where all the age groups behave similarly, displaying equivalent response patterns to the same tasks, and all of those tasks coincide in unequivocally displaying equivalent inhibitory and monitoring abilities between monolinguals and monolinguals, it is quite unlikely that age variations account for this null results. If that were the case, considering that the bilingual advantage hypothesis was tested in various age groups, bilinguals should have had performed better in some age groups and equivalently to monolinguals in others. However, they all behaved the same.

b. Previous significant results were a consequence of unmatched and small samples

If age variations of different cognitive skills are not ultimately responsible for the absence of the bilingual advantage, then the first coherent explanation for the so-called bilingual advantage to appear in some of the studies reported above are the methodological issues explained in the first chapter (Paap & Greenberg, 2013) regarding the affecting external factors and sample sizes. In this thesis I attempted to verifying the reliability of the bilingual advantage hypothesis by using the same tasks and equivalent populations as previous studies (Bialystok & Martin, 2004, Bialystok & Shapero, 2005; Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012; Bialystok, Craik, & Luk, 2008) whilst bearing in mind all the concerns raised by the skeptical side of the debate (Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015a, 2015b) regarding sample sizes, participant selection and group matching, to be able to account for them. I deal with this issue in the present thesis, given that the sample sizes were clearly large (especially in children and young adults), representing the largest sample tested so far in this regard. Besides, the matching of the groups was done taking into account the age, general IQ test immigrant status and SES (see Methods section of each chapter), which has been argued to be directly tied to advantages in executive functions. I also tried to eliminate any potential influence of immigrant status by restricting the inclusion of participants to those who lived in (and were originally from) the same country. Hence, to my eyes, the only relevant and evident difference between groups corresponded to their linguistic profile. When the bilingual advantage is tested in an extensive study with carefully matched samples of elderly, children and young adults, the alleged advantage completely disappears. These results favour the perspective of the skeptical side of the bilingual advantage debate, and provide credibility to the concerns that the bilingual advantage obtained previously in these tasks might be due to uncontrolled external factors (Paap & Greenberg, 2013) and that it disappears when the spurious factors are controlled for (Paap, Johnson, & Sawi, 2015a, 2015b; Gathercole et al., 2014). These consistently increasing null findings should be considered meaningful and representative of the general population, despite the well-known publication bias towards the evidence showing a bilingual advantage and against publishing the challenging its existence (see de Bruin, Treccani, & Della Sala, 2014; for a revision in the mentioned publication-bias).

c. *No bilingual advantage at the behavioral level: Bilingualism is not enough to create differences*

The growing body of evidence showing no bilingual advantage seems to indicate that the reported significant results are a consequence of non-rigorous methodological praxis (i.e., uncontrolled external factors). Therefore, one could conclude that if all the previously shown bilingual advantages are not a consequence of speaking two languages, then the alleged extra training that the bilingualism provides with in terms of inhibition and monitoring is not strong enough to provoke changes at the behavioral level. Namely, the argument for a bilingual advantage on executive control tasks rests on the idea that monolinguals do not switch between two languages, since they only have one available. In comparison, bilinguals do and that is where benefits should come from. However, all human beings face situations in which they have to inhibit salient responses constantly and monitor the environment, in both general social situations and when performing concrete actions. For example, people do switch between comprehension and production when they talk to somebody, they do switch and keep their monitoring abilities strongly activated when they have to drive and talk to somebody, or they inhibit salient responses when they have to adapt their speech and manners to different social situations, that can range from casual to very formal. Thus, monolinguals also efficiently use their switching, inhibitory and monitoring skills in many other domains, and it is unclear whether language switching in bilinguals - although much more present than in monolinguals - imposes a heavier burden than the one imposed to everyone, monolingual or bilingual, in their daily life.

The first proposal of these conclusions is that bilingualism does not enhance general executive functioning to the point of being overtly better than monolinguals' at the behavioral level, and that when it was found it was due to methodological problems such as poorly matched groups. Interpreting this data from the classic *modularity model* (Fodor, 1983), it could have been claimed that even if bilinguals make use of the general inhibitory abilities more often when applied to language processing, it is a very concrete aspect to which inhibition can be applied. This ability would be part of the central system and thus domain general, so the specific use of it in language would not enhance it when compared to monolinguals because monolinguals would also use it for dozens of other actions that would, at the end of the day, make it as efficient and as enhanced as bilinguals'. On this perspective, bilingualism would not impose such important training.

d. *Executive functions are not domain general: bilingualism only enhances language control*

These results can also be explained from an alternative perspective, coming from data very recently published. Bilinguals surely have between-language-connections, as demonstrated by many tasks that show a modulation of phenomena such translation or cognate effects (Costa, 2005; Costa, Santesteban, & Caño, 2005). As a consequence, they have to apply inhibition to their non-target language, as seen by many studies showing longer reaction times in language switches (Thomas & Allport, 2000). Intuitively thinking, they should be better than monolinguals at applying this inhibition, by training transfer, as they do it more often when they deal with languages. Why is it absent in domain-general tasks that require inhibiting irrelevant information? One possibility is because this enhancement would only be present when inhibitory abilities are required in linguistic contexts. Because, as defended by recent publications, executive functions might be domain specific and *not* domain general.

As explained in Chapter 1, how bilinguals manage their two languages to restrict their activation and prevent massive influence and interference between them has been a very attractive topic in the field in the last several years (Costa & Santestban, 2004; Jackson, Swainson, Cunnington, & Jackson, 2001; Branzi, Della Rosa, Canini, Costa, & Abutalebi, 2015; and see Baus, Branzi, & Costa, 2015; for review). The field agrees on the fact that bilinguals make use of domain general executive control functions to control their languages (Abutalebi & Green, 2007). However, the defenders of the bilingual advantage have taken for granted that language control and general executive control functions completely overlap or that the improvement in one directly implies improvement in the other (e.g. "Crucially, the mechanisms that reduces attention to the non-relevant language system is the same as that used to manage attention in all cognitive tasks", Bialystok et al., 2005, p.41), but such a statement is not self-evident. One could intuitively parallel this to Fodor's modularity model (1983), and assume that the inhibition (applied to language or not) should be considered a high level skill that is part of the central system, and therefore it would extend its effect to any input received by the system, linguistic or not. Thus, it makes sense to assume that, as long as it is a single component that applies to anything that needs to be inhibited, its improvement via language use (if it happens) would have been reflected in any other inhibition related task. However, the recent findings that will be explained in the upcoming paragraphs seem to suggest that inhibition can be domain specific and that different kind of inhibition would respond distinctively to different situations or stimuli types, thus suggesting that each input module, which are domain specific in Fodor's model (1983), would require its own inhibitory function and thus the improvement of linguistic inhibition would not necessarily mean the improvement of inhibition skills applied to other domains.

Different efforts have been made in different venues to try to explore this possibility. For example, one of the most extended approaches to try to capture any overlap between language control and general executive functions have been to test participants in both domain general executive function tasks and language control tasks and correlate their behavior among tasks (Calabria, Hernández, Branzi, & Costa, 2012; Prior & Gollan, 2013; Calabria, Branzi, Marne, Hernández, & Costa, 2015). If both tasks tap into comparable cognitive abilities responsible of their behavior in all of the tasks, correlations should be found. For example, Calabria et al. (2012; 2015) tested bilinguals of different ages in tasks tapping into shifting by using the n-1 cost paradigm (Kiesel et al., 2010) in a linguistic and non-linguistic version. For the linguistic version, participants had to name pictures in either their L1 or their L2, and the cost of switching from one language to another in the next trial (switch trials, L1 after naming L2 or vice versa) was compared to the reaction times of trials with no language switch (repeat trials, L1 after naming in L1 or L2 after L2). For the non-linguistic version, participants had to classify pictures based on shape or colour, depending on the cue presented together with the picture. Similarly, reaction times to items that required a change in the classification criteria (switch trials, e.g. responding based on colour and then responding based on shape) were compared to the ones in which the criteria did not change (repeat trials, e.g., two colour judgements). Once linguistic and non-linguistic switch costs were obtained, correlations were performed. There was no correlation between linguistic and non-linguistic n-1 switch costs (see also Prior & Gollan, 2013), suggesting that the underlying constructs responsible for the performance in tasks involving language control and domain general control are not the same. Opposing to this view, some other studies have found relation between the frequency rate in which bilinguals would switch between their languages in a daily basis and the mixing costs (in error rates) in a set-shifting task (Soveri, Rodriguez-Fornells, & Laine, 2011). Comparably, other findings report a link between cognitive measures of executive functioning and intrusion error rates in single-language conversational settings (Festman, 2012; Gollan, Sandoval, & Salmon, 2011). Closer to the theory of the bilingual advantage, studies that focused on the relation between language control and non-verbal interference control show inconclusive results. For example, Prior and Gollan (2011) reported that Mandarin-English bilinguals showed a negative correlation between their fluency scores in Mandarin and the switch costs sizes in a non-linguistic switching task, but Spanish-English bilinguals tested in the same study with the same parameters showed no replication of the finding.

Branzi and her colleagues (Branzi, Calabria, Boscarino, & Costa, 2016) defend that, in order to disentangle this debate, the overlap between general executive process tasks and language control tasks should indeed be considered as one of the most informative sources. However, they argue, the correlations reported previously between the n-1 linguistic and non-linguistic tasks (Calabria et al., 2012; 2015) might have used tasks that do not tackle specifically in the skills that bilingualism is argued to influence, i.e. the inhibitory abilities (Green, 1998). N-1 tasks not only use inhibitory control, but also other EC mechanisms that are more involved in switching processes (Kiesel et al., 2010). One could argue that the lack of correlation between linguistic n-1 and non-linguistic n-1 does not necessarily imply that the same processes are not involved in both of them, but maybe the variability of the other EF aspects involved could camouflage the existing direct correlation between inhibition involved in linguistic and non-linguistic switching paradigms. A solution for that dilemma is the n-2 repetition cost (Mayr & Keele, 2000; Philipp, Gade, & Koch, 2007). N-1 defines the trials in comparison to the previous one, whether a switch happened or not. In contrast, n-2 defines the trials in comparison to the second to last trial, and its repetition cost refers to the reported slower RTs when participants have to switch into recently performed task (which was inhibited to be able to perform the upcoming demands) as compared to when participants have to perform a task that was not recently demanded (and therefore not inhibited). For example, when participants have to classify pictures based on colour, then size, and then colour again, they are slower than when they have to classify pictures based on shape, size and then colour, because the first *schema,* responsible of the colour task, was inhibited in the first series. This difference is obtained when RTs for the last colour item of both series are measured, and it is called the n-2 repetition cost, argued to reflect inhibition applied to the repeated task.

Branzi and colleagues (2016) explored this same paradigm in 62 early and highly proficient Catalan-Spanish bilinguals, who performed linguistic and non-linguistic n-2 tasks as well as linguistic and non-linguistic n-1 tasks. If language control makes use of the same inhibitory control mechanisms as domain general executive control, then the two tasks should show a correlation between the linguistic and the domain general variety (n-1 for switching and n-2 for inhibition), and thus the bilingual advantage would have a realistic grounding. For the linguistic n-2 task, participants were required to name out loud pictures in their L1, L2 or L3 in settings that would create repeated trials (ABA sequences, e.g., L1, L2, L1) or not-repeating trials (CBA, e.g. L3, L2, L1). In the non-linguistic task, the same procedure was applied to visual stimuli that needed to be classified based on different physical features (similarly to Philipp, & Koch, 2006). After meticulous analyses, authors showed evidence that indicated that, despite results of n-1 and n-2 tasks were correlated (indicating a presence of a common EC mechanism in the two of them), the effect of the linguistic n-2 repetition cost was smaller than that of n-1 repetition's and, importantly, neither n-1 shift costs nor n-2 repetition costs were not correlated across linguistic and non-linguistic tasks. In other words, the participants' performance on non-linguistic tasks and linguistic tasks was completely unrelated. If the n-2 repetition cost, which is a reliable index of inhibitory control (Mayr, & Keele, 2000), does not correlate across tasks when the stimuli used are different (see also different n-2 patterns depending on the language used in Babcock & Vallesi, 2015), the underlying inhibitory mechanisms applied to linguistic and non-linguistic situations could, arguably, be different.

Those results, which in principle might surprisingly suggest that the executive functioning abilities (and mainly inhibition) required in linguistic and non-linguistic tasks are not the same, should not be unexpected. Even neuroimaging studies have suggested that there is not a complete overlap between domain general executive functions and language specific ones (Magezi, Khateb, Mouthon, Spierer, & Annoni, 2012; Branzi et al., 2015). Furthermore, even in the very same field of language control, the most recent studies speak of different language control mechanisms relying on different neural substrates when applied to language comprehension and production (Blanco-Elorrieta, & Pylkkänen, 2016).

Therefore, if recent research conducted in highly proficient native balanced bilinguals (equivalent to the samples used here) speak of different control mechanisms and inhibitory capacities that are responsible for language control and domain general tasks, why should anyone expect an enhancement of one of them when the other is trained? If even language control procedures seem to be different in production and comprehension processes, why should bilingualism enhance something completely unrelated from language such as general inhibitory control? I consider that the results exposed in the previous paragraphs account for the null results reported in all the experiments that I presented in this thesis as well as other researchers in the field that argue for a lack of enhanced general inhibitory abilities in bilinguals as a consequence of a better trained language control (Paap & Greenberg, 2013; von Bastian, Souza & Gade, 2016; Paap, Johnson, & Sawi, 2015). As a clear example of the domain specificity of inhibition, the potential training transfer of this ability has been tested in children by training them in domain-general inhibitory and working memory tasks, and testing them later in the same trained and new untrained tasks (Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2008). The authors found that training improved the performance in all the trained tasks. Crucially, while the non-trained memory tasks also showed a benefit from training transfer, no improvement was found in the non-trained inhibitory tasks.

As the second proposal that would account for the data on the present thesis, and based on recent data (Branzi et al., 2016), here I argue that the executive functions (and concretely inhibition) that are required in language control are not necessarily the same as the ones required in domain general tasks, and thus the bilingual training would only enhance language control and not domain general executive functions. These conclusions are strongly reinforced by the findings, reported at the end of the “results” section of each chapter, indicating that the correlation between task indices are null or, if significant, inverse, indicating that even the indices of inhibition across similar tasks (but using different stimuli) don not correlate with each other. Similarly, consider here also recent reviews arguing for this same cross-task low replicability and reliability (Paap & Greenberg, 2013, Paap & Sawi, 2014). One might argue that expecting to find strong advantages in non-linguistic tasks as a consequence of an enhancement of language specific control mechanisms (as tested by Branzi et al., 2016) lacks of any coherent grounding. Especially, considering that they seem to measure different particular inhibitory processes, not related to each other, as shown by the low cross-task correlations.

III. Possible situations in which an advantage might be found

Notwithstanding the results presented in the current thesis, it might be worth considering that a bilingual advantage could arise in different kinds of populations. Such an effect, that is present in some studies and absent in others, might not be completely ungrounded and they might actually arise in some concrete conditions.

a. Immigration

In this regard, many have been the researchers that pointed out the importance of social factors when bilingualism is explored (Reynolds, 1991), especially when applied to group comparisons exploring the bilingual advantage (Paap & Greenberg, 2013). As pointed out by Morton & Harper (2007), one of the most determining factors has been the issue of the immigrant status. Revision of the literature reveals that many (if not most) of the research reporting bilingual advantages included immigrants as the majority of their bilingual samples and non-immigrant in the monolingual one (see Bialystok, 1986; Bialystok & Senman, 2004; Bialystok & Shapero, 2005; Bialystok, 1999 and Bialystok & Martin, 2004; among others), which is especially obvious in studies that test older bilinguals and monolinguals in executive functioning or in the effects on the onset of dementia (Bialystok, Craik, & Luk, 2008; Bialystok et al, 2008; Gold, Kim, Johnson, Kryscio, & Smith, 2013; Salvatierra & Roseselli, 2010; Schroeder & Marian, 2012; Bialystok, Craik, & Freedman, 2007). Crucially, when immigrant status is controlled for, group differences tend to not occur (Kirk et al, 2014; Kousaie & Phillips, 2012b; Morton & Harper, 2007, see Cherktow et al., 2010; for benefits for bilingual immigrant samples but not for bilingual local samples) and importantly, when the hypothesis of bilingualism delaying symptoms of dementia is tested in prospective cohort studies, no benefit of bilingualism is found (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).

These data seem to suggest that being an immigrant, and not bilingualism, is providing individuals with some benefits. Upon review, research shows that, once SES is adjusted for, immigrants have better morbidity and mortality outcomes than non-immigrants in different parts around the world (Crimmins, Soldo, Kim, & Alley, 2005; Thomson, Nuru-Jeter, Richardson, Raza, & Minkler, 2013; Palloni & Arias, 2004; Ng, 2011; Strong, Tricket, & Bhatia, 1998; Kreft & Doblhammer, 2012), which is known as the “healthy immigrant effect”. This led some researchers to argue that maybe the reality selects the healthiest individuals to move around, and they are the ones that decide to travel and able to pass health screening tests in the host countries (Kennedy, McDonald, & Biddle, 2006). Not only they are healthier, but even when cognitive functioning is checked after adjustment for English proficiency, financial situation, education, age, physical health and health behavior, immigrants show slower decline rates than host-country born individuals (Hill, Angel, Balistreri, & Herrera, 2012). One might argue that not only immigrants might have shown advantages because they are healthier, but also because they show higher IQ. Due to the migration policies of countries such Canada and Australia, countries seek immigrants with high levels of education and skills, and even with job offers. Due to that fact, a diverse variety of studies have found higher IQ values for people who left their homeland as compared to their peers who stayed home (Milne, Poulton, Caspi, & Moffitt, 2001; Kuhn, Everet, & Silvey, 2011; Wadsworth, Kuh, Richards & Hardy, 2006). Given that IQ has been shown to be highly protective against dementia and cognitive declining (Fritsch et al., 2005), immigrants would be more protected as compared to non-immigrants.

All this evidence has led some researchers to consider the healthy migrant effect as the confounding link between bilingualism and the delayed onset of the symptoms of dementia. Immigrants, who by force of their reality are forced to be bilinguals, feature advantages and enjoy some protection when they are compared to non-immigrants, who happen to be monolinguals.

Supporting this hypothesis, and as commented in the previous paragraphs, when the bilingual hypothesis has been tested with non-immigrant samples (and of course controlling many other confounding sociodemographic variables), any trace of bilingual advantage has vanished in children, young adults and the elderly. Along the same lines, the results exposed in the present thesis coming from experiments testing non-immigrant bilinguals and monolinguals from the same area, and also studies showing no differences between monolinguals and non-immigrant bilinguals tested in bilingual areas such as Wales (de Bruin, Bak, & Della Sala, 2015), seem to suggest that bilingualism without immigration does not enhance any cognitive skills.

b. Nature vs. nurture

The effect of immigration and its relation with bilingualism and the eventual enhancement of the executive functions could be interpreted from a different perspective: the need to migrate could bring some cognitive benefits because it forces individuals to adapt to new social environments that demand mastery of a second language from these newcomers, eventually making them bilinguals. It has been argued that, if bilingualism has any enhancing effect, it should be easily perceived in native bilinguals when compared to late bilinguals (Luk, De Sa, & Bialystok, 2011). However, other studies have suggested that the effects might be clearer in late bilinguals (Tao, Marzecová, Taft, Asanowic, & Wodniecka, 2011; Vega-Mendoza, West, Sorace, & Bak, 2015, for review).

This theory might well explain the results from the studies I presented in this thesis. It might be coherent to propose that native bilingualism does not necessarily lead to an eventual benefit because there is no cognitive effort of dealing with two languages since birth. It might be worth suggesting that bilingual natives who deal with two languages since birth can easily incorporate any language (i.e., irrespective of being one or two) into their repertoire as long as they are given early in life. In those individuals, no reconfiguration would be needed, they are born with those languages and it is their default state to be a bilingual. Here I propose an alternative approach to explore the potential benefits of bilingualism in executive control: late acquired bilingualism would indeed require new bilingual speakers to re-adjust their mental repertoires and accommodate the existing system to the newly acquired language. This,

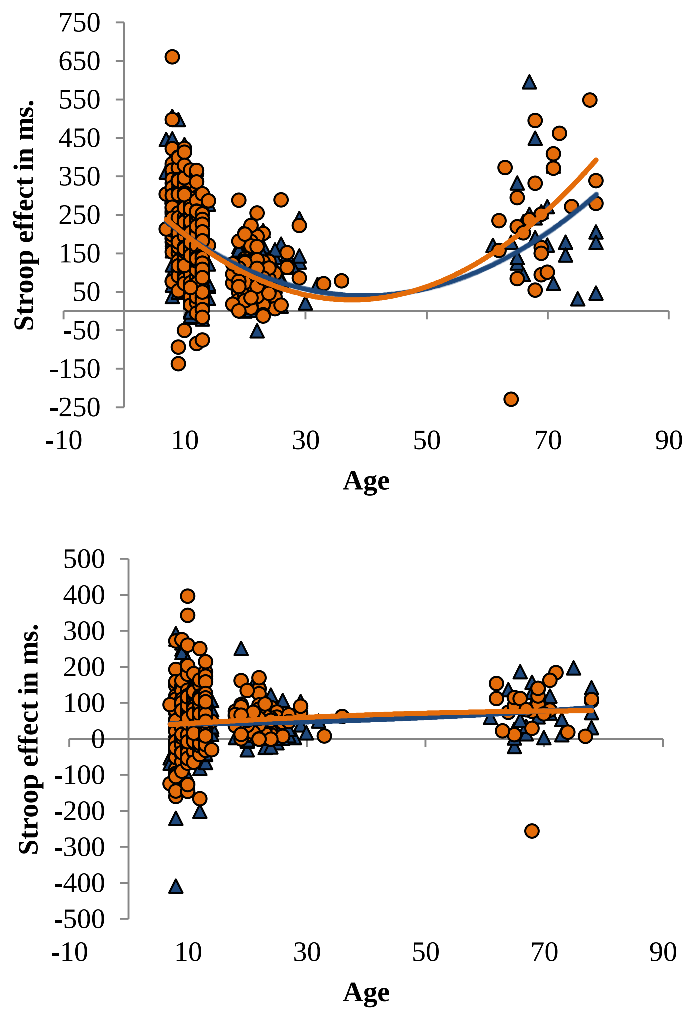

Figure 22. Developmental trends of the Stroop indices across age. Stoop indices obtained in the verbal (upper panel) and numerical (lower panel) tasks are plotted in relation to the age of the participants. Orange dots and lines represent monolingual participants, and blue triangles and lines represent bilingual participants.

inasmuch as it requires a cognitive effort to adapt the system, could lead to an improvement, in the same way that training and cognitively demanding acquired skills lead to an enhancement of attentional skills (Dye, Green, & Bavelier, 2009a;b). It is worth mentioning that when the transfer of the training to concrete cognitive abilities have been argued to be stronger is when the new skills are acquired later in life (Hillman, Erickson, & Kramer, 2008).

Finally, bringing the "Nature vs. Nurture" debate (Pinker, 2003; Meaney, 2004; Powledge, 2011; Ridley 2003; Normile, 2016) to the bilingual advantage field, I deem bilingualism able to shape individuals general cognitive skills when it is faced in an effortful fashion late in life (i.e., nurture) and not when it is given by default to a bilingual-born baby (i.e., nature) who may even start dealing with two languages prenatally (Byers-Heinlein, Burns, & Werker, 2010; May, Byers-Heinlein, Gervain, & Werker, 2011).

Despite the absence of direct evidence in this regard, but based on the evidence shown through the different chapters and conclusions of the present thesis, I argue that if bilingualism should provide bilinguals with any relevant cognitive benefit, they are not present in balanced, native and non-immigrant bilinguals. Indeed, if we take a developmental perspective of the data on the current thesis by computing the Stroop indices of both the verbal and numerical Stroop tasks (the only tasks that were present in all the age groups) and we observe how it develops through the lifespan (see Fig. 22), it is obvious that bilingualism did not play any important role in its development. The trends presented there clearly overlap with each other[8].It seems that, when administrated since birth (and even earlier), bilingualism seems not to modulate any development of the inhibitory control development as measured by the Stroop tasks presented here. Instead, they should be captured (if there is any effect) in bilinguals that acquired their second language late in life (Antoniou, Gunasekera, & Wong, 2013) which, very likely, could happen as a consequence of immigration. Learning a new language while (or because of) facing the high demand of being an immigrant in a foreign country could boost the mental circuitry to the limit and create what has been misleadingly

[8] Two series of regression models were created to try to explain the trend of the Stroop effect size in each task. The model only including Age was tested against the other one including Age and Group (monolinguals and bilinguals), using "anova" function in "lme4" package in R software (Bates, 2013; Team, 2014; version 3.0.2). In both cases the differences between the more complex and the more simple model were non-significant (all *F*s<1, all *p*s>.54), which indicates that adding context does not improve the explanatory power and that, therefore, the simpler model should be adopted. Although the explanatory power of the model containing only age is not big, showing R2 values of around .2 for the verbal Stroop task (.20 for bilinguals and .25 for monolinguals) and around .02 in the numerical Stroop (similar in both groups), adding the factor of language context doesn't improve it.

named the bilingual advantage. 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 for a study showing a positive correlation between the years of immersion in an L2 context and the size of the *Stroop effect*), because that would be the moment of the most demanding adaptation process. In that moment, if we had an equivalent graph to the ones showed above, the inclusion of a second language in individuals' mental repertoire could change the trend of the developmental tendency of the executive functions.

This would explain some results observed in the literature and in the present thesis, given the characteristics (natives, balanced, lifelong bilinguals) of the populations tested in the current work. For future research, and 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 later in life, in the same group of individuals, following a longitudinal approach. It is still an open question whether or not a bilingualism acquired later in life, in the middle of the aging process showed in the Fig. 22, would alter the way each of the groups' trend eventually develop.

Conclusions

The present thesis aimed at exploring the existence of the bilingual advantage in executive functions throughout the whole lifespan. It has been argued that bilingualism requires very efficient use of executive control abilities to properly manage two languages and avoid unwanted cross-language effects such as intrusions. This would eventually enhance the executive function abilities required in domain general situations (i.e. not necessarily language-related) by training transfer. Concretely, bilingualism should enhance inhibition (stemming from the constant need of inhibiting the non-target language) and monitoring (which is produced by the constant checking of the needs of the environment and interlocutors to be able to quickly adapt to the demanding changes). In contrast to several findings in the literature that advocate for such an enhancement, several recent studies have criticized the so-called bilingual advantage by arguing that it is actually a consequence of non-rigorous experimental praxis and a product of uncontrolled factors that would make bilingual and monolingual groups under study to differ in several relevant factors apart from linguistic profile. This skeptical side argues that the bilingual advantage that is found in the literature is not a consequence of bilingualism, but of other factors that are unevenly distributed across samples of bilinguals and monolinguals and, when left uncontrolled, create different patterns of response that have been wrongly associated to the so-called bilingual advantage.

In the present thesis I tested the demographic groups that were most likely to show an advantage (i.e., elderly and children) together with the most studied demographic target (young adults) in a variety of tasks previously used in the literature showing results supporting the bilingual advantage. Furthermore, the sample sizes here were much larger than in the studies showing an advantage, and the external relevant factors were carefully matched. None of the tasks conducted in any of the demographic groups yielded significance in any of the critical indices that linguistic groups were compared in. 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. The bilingual advantage was not found in senior samples either, and thus the argument for a lifelong experience boosting cognitive abilities loses credibility. Finally, young adults were tested in four classic psychological tasks that tap into inhibitory abilities, and none of them yielded significant results.

Thus, the evidence collected in the present thesis 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, meaning that the latter do not outperform the former in inhibitory abilities and neither do they in monitoring skills.

Different perspectives are taken to try to explain the null findings obtained in the current work. Firstly, it is argued that, if executive functions are domain general (as defended, for example, by Bialystok et al., 2005), and the same abilities are applied to linguistic situations (i.e., the aspect in which bilinguals would be trained much more than monolinguals) as well as other general contexts in which they are required (i.e., daily life contexts in which both bilinguals and monolinguals would need to use those abilities), bilingualism would not play such an important role in the development of the executive functions. If those abilities are applied to any situation in which inhibition and monitoring are required, the extensive use of them that both bilinguals and monolinguals feature every day would make the impact of bilingualism irrelevant, given that language context would be just a small percentage of the situations in which executive functions are needed. The first proposal is that bilingualism would not suppose such a strong burden to create such an important boosting, because is just another aspect of life (among dozens) in which executive functions have to be used, and thus this extra training would not be that relevant. This perspective is based on recent data that suggests that brain circuitry involved in certain executive functions, like language switching, are the same as those involved in general switching tasks (see, for example, de Baene, Duyck, Brass, & Carreiras, 2015) and it accounts for the increasing studies that show equivalent performance across groups.

Secondly, and opposed to the first hypothesis explained in the previous paragraph, I argue in favour of different executive control abilities applied to different kind of situations. Thus, executive functions would be domain specific, and therefore training in language-related executive functions (produced by bilingualism) should not necessarily imply an enhancement on the executive control abilities when applied to non-linguistic situations. Any enhancement, if present, would be only applied to that field in which it has been trained, with no training transfer to any other aspect of cognition. This perspective is based on recent evidence that shows that behavioral results on domain general inhibition based tasks and language control tasks that also require inhibition do not correlate with each other. Supporting this hypothesis, I found non-significant correlations between the indices of the different tasks used in the present thesis, arguably indicating different underlying mechanisms. Also, if language control indeed relied on domain general EF mechanisms that are used in any kind of situation, one would expect that the better somebody is in the use of their L2 and in the managing of two or more languages, the better their performance in non-linguistic tasks that require EF abilities would be. However, no modulation of said performance, based on the tested participants' L2 proficiency, was found in the present thesis.

Thirdly, since I only tested native bilinguals who acquired both languages very early in life, I acknowledge that the possibility of bilingualism enhancing domain general executive abilities when it is acquired later in life should not be dismissed. This idea is based on evidence that shows a bilingual advantage mainly when the bilingual group under test is formed of immigrants, and also on evidence showing cognitive advantages in immigrant groups. It could be the case that only the smartest and cognitively healthiest become immigrants, but on the other hand, becoming a bilingual because of immigration comes from the fact of changing one's country and language later in life, so this strong readjustment to language and context might be the cause of the advantages found. In order to disentangle whether it is the readjustment to a new language or to a new context what creates the advantage, this question needs to be addressed by future research. I hereby advocate for stronger changes produced by lately acquired second language than by two languages provided from birth, since in the latter case neither especial effort nor readjustment is required and thus one would not expect the special boosting of any ability. However, readjusting the whole cognitive system to a newly acquired language, which is a complex system, and to a new context (in the case of immigrants) might have an impact that is still not completely comprehended.

In a nutshell, the conclusion of this thesis is that native, balanced and lifelong bilingualism does not enhance bilinguals' executive functions -concretely inhibition or monitoring- when compared to appropriately matched monolinguals in studies using domain general tasks in experiments with large sample sizes.

But extreme caution is advised in the interpretation of the present thesis. By no means is my intention to propose that bilingualism does not have any benefit for bilingual speakers. Speaking more than one language is a wonderful tool to see the world with different eyes, be open to other people and other cultures, and it definitely does provide individuals with some benefits that completely balance out the small disadvantages that are mainly found in experimental settings (see Chapter 1 for a more detailed explanation of advantages and disadvantages provided by bilingualism). The argument of this thesis is also restricted to concrete situations in experimental settings in which the impact of bilingualism on particular psychological constructs has been under debate, and emphasizes the importance of good experimental praxis in order to avoid incorrect or incomplete generalizations. Other than that, the knowledge of two (or more) languages is something that should be strongly encouraged, reinforced and promoted by private and public entities, because its positive effects are numerous and the small and insignificant negative impacts are almost never noticeable in daily life; but largely restricted to scientific settings.

List of publications derived from the data presented on this thesis

Duñabeitia, J.A., Hernández-Cabrera, 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. DOI: 10.1027/1618-3169/a000243.

Antón, E., Duñabeitia, J.A., Estévez, A., Hernández-Cabrera, 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, Vol. 5*, 398, 1-12. DOI: 10.3389/fpsyg.2014.00398

Antón, E., Fernández García, Y., Carreiras, M., & Duñabeitia, J.A., (2016) Does bilingualism shape inhibitory control in the elderly? *Journal of Memory & Language, 90*, 147–160. DOI: 10.1016/j.jml.2016.04.007

Antón, E., Carreiras, M., & Duñabeitia, J.A., (2016). On the effects of bilingualism on executive functions and working memory in young adults. Manuscript submitted for publication.

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta psychologica, 128*(3), 466-478.

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of neurolinguistics, 20*(3), 242-275.

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and cognitive processes, 23*(4), 557-582.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., ... & Costa, A. (2012). Bilingualism Tunes the Anterior Cingulate Cortex for Conflict Monitoring. *Cerebral Cortex, 22*(9), 2076-2086.

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

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from a new paradigm. *Developmental Science, 14*, 395–401

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence, 8*(2), 71–82.

Antoniou, M., Gunasekera, G. M., & Wong, P. C. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: a hypothesis for future research. *Neuroscience & Biobehavioral Reviews, 37*(10), 2689-2698.

Arbib, M. (1987). *Modularity and interaction of brain regions underlying visuomotor coordination*. In J. L. Garfield (ed.), Modularity in Knowledge Representation and Natural-Language Understanding, Cambridge, MA: MIT Press, pp. 333–363.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology, 5*(2), 115-129.

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

Babcock, L., & Vallesi, A. (2015). Language control is not a one-size-fits-all languages process: evidence from simultaneous interpretation students and the n-2 repetition cost. *Frontiers in psychology, 6*.

Baker, C., & Jones, S. P. (Eds.). (1998). *Encyclopedia of bilingualism and bilingual education*. Multilingual Matters.

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*(1), 51–64.

Bates, D., Maechler, M., Bolker, B., Walker, S., (2013). lme4: Linear mixed-effects models using Eigen and S4. *R package version 1* (7) 0-4.

Baus, C., Branzi F. M., & Costa A. (2015). *On the mechanism and scope of language control in bilingual speech production.* The Cambridge Handbook of bilingual processing.

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of memory and language, 26*(6), 658-672.

Ben-Zeev, S. (1977). The Influence of Bilingualism on Cognitive Strategy and Cognitive Development. *Child Development, 48(*3), 1009-1018.

Berg, T., & Schade, U. (1992). The role of inhibition in a spreading-activation model of language production. I. The psycholinguistic perspective. *Journal of Psycholinguistic Research, 21*(6), 405-434.

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

Beveridge, M., Jarrold, C., & Pettit, E. (2002). An experimental approach to executive fingerprinting in young children. *Infant and Child Development, 11*(2), 107-123.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development, 57*, 498–510

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child development, 70*(3), 636-644.

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. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*(01), 3–11.

Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Canadian Journal of Experimental Psychology, 65*(4), 229.

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 565.

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition, 15*(02), 397-401.

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., & Senman, L. (2004). Executive processes in appearance–reality tasks: The role of inhibition of attention and symbolic representation. *Child development, 75*(2), 562-579.

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

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., Barac, R., Blaye, A., & Poulin-Dubois, D. (2010). Word Mapping and Executive Functioning in Young Monolingual and Bilingual Children. J*ournal of Cognition and Development: Official Journal of the Cognitive Development Society, 11(*4), 485–508.

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia, 45(*2), 459–464.

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., 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., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24*(1), 40-49.

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., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(04), 525-531.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics, 24*(01), 27-44.

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

Bishop, D. V. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. *The Quarterly Journal of Experimental Psychology: Section A, 50*(4), 899-923.

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual Language Control in Perception versus Action: MEG Reveals Comprehension Control Mechanisms in Anterior Cingulate Cortex and Domain-General Control of Production in Dorsolateral Prefrontal Cortex. *The Journal of Neuroscience, 36*(2), 290-301.

Block, N. (1995). The mind as the software of the brain. In E. E. Smith & D. N. Osherson (Eds.), *Thinking: An invitation to the cognitive science* (pp. 377-425). Cambridge, MA: MIT Press.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition, 17*(03), 610-629.

Bock, K., and Levelt, W.J.M. (1994). Language production. Grammatical encoding. IN M.A. Gernsbacher (Ed.). *Handbook of psycholinguistics* (pp.741-779). New York: Academic Press

Boersma, P & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 6.0.05, retrieved 8 November 2015 from http://www.praat.org/

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2,* 29–49

Bradley, K. A., King, K. E., & Hernandez, A. E. (2013). Language experience differentiates prefrontal and subcortical activation of the cognitive control network in novel word learning. *NeuroImage, 67,* 101-110.

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

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2015). Language Control in Bilinguals: Monitoring and Response Selection. *Cerebral Cortex, 1,* 14.

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental neuropsychology, 26*(2), 571-593.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?. *Bilingualism: Language and cognition, 13*(03), 359-371.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The Roots of Bilingualism in Newborns. *Psychological Science, 21*(3), 343-348.

Calabria, M., Branzi, F. M., Marne, P., Hernández, M., & Costa, A. (2015). Age-related effects over bilingual language control and executive control. *Bilingualism: Language and Cognition, 18*(01), 65-78.

Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2012). Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology, 2*, 399.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive neuropsychology, 14*(1), 177-208.

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex, 26*(1), 95-122.

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental science, 11*(2), 282–298.

Carruthers, P. (2006). *The Architecture of the Mind*, Oxford: Oxford University Press.

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*(4), 264–272.

Cenoz, J., & Valencia, J. F. (1994). Additive trilingualism: Evidence from the Basque country. *Applied Psycholinguistics, 15*(02), 195-207.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. (2001). Changes in executive control across the life span: examination of task-switching performance. *Developmental psychology, 37*(5), 715.

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

Churchland, P. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science, 55:* 167–187

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.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior, 8*(2), 240-247

Colomé, À., & Miozzo, M. (2010). Which words are activated during bilingual word production*? Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 96.

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*(2), 302–312.

Cook, V. (1997). The consequences of bilingualism for cognitive processing.*Tutorials in bilingualism: Psycholinguistic perspectives,* 279-299.

Costa A. (2005). Lexical Access in Bilingual Production. In Kroll, J. F. & De Groot, A. M.B. (Eds.) *Handbook of Bilingualism: Psycholinguistic Approaches.* Oxford University Press.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition, 2*(03), 231-244.

Costa, A., & Santesteban, M. (2004a). Bilingual word perception and production: two sides of the same coin? *Trends in cognitive sciences, 8*(6), 253-253.

Costa, A., & Santesteban, M. (2004b). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of memory and Language, 50*(4), 491-511.

Costa, A., & Santesteban, M. (2006). The control of speech production by bilingual speakers: Introductory remarks. *Bilingualism: Language and Cognition, 9*(02), 115-117.

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

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283.

Costa, A., Caramazza, A., & Sebastián-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283.

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., 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., La Heij, W., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition, 9*(02), 137-151.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41*(3), 365-397.

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and language, 94*(1), 94-103.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in cognitive sciences, 10*(3), 131–138.

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

Crimmins, E. M., Soldo, B. J., Ki Kim, J., & Alley, D. E. (2005). Using anthropometric indicators for Mexicans in the United States and Mexico to understand the selection of migrants and the "Hispanic paradox". *Social Biology, 52*(3-4), 164-177.

Crystal, D. (1997). *English as a global language*. Cambridge, England: Cambridge University Press

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science,320*(5882), 1510-1512.

Darcy, N. T. (1946). The effect of bilingualism upon the measurement of the intelligence of children of preschool age. *Journal of Educational Psychology,37*(1), 21.

Darcy, N. T. (1953). A review of the literature on the effects of bilingualism upon the measurement of intelligence. *The Pedagogical Seminary and Journal of Genetic Psychology, 82*(1), 21-57.

Darcy, N. T. (1963). Bilingualism and the measurement of intelligence: Review of a decade of research. *The Journal of genetic psychology, 103*(2), 259-282.

De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain circuit for cognitive control is shared by task and language switching. Journal of Cognitive Neuroscience, 27(9), 1752–1765.

De Bot, K. (1992). A Bilingual Production Model: Levelt's “Speaking” model adapted. *Applied Linguistics,13*(1), 1-24.

De Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. *The bilingual lexicon, 191*, 214.

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

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism an example of publication bias? *Psychological science, 26*(1), 99-107.

De Groot, A. M. (1995). Determinants of bilingual lexicosemantic organisation. *Computer Assisted Language Learning, 8*(2-3), 151-180.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1-2), 1-42.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General,122*(3), 371.

Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2010). Executive functions in decision making: An individual differences approach. *Thinking & Reasoning,16*(2), 69-97.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review, 93*(3), 283.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991). *Psychological review, 98(*4), 604-614.

Dell, G. S., & O'Seaghdha, P. G. (1994). *Inhibition in interactive activation models of linguistic selection and sequencing.* Academic Press.

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science, 23*(4), 517-542.

Deuchar, M. (2005). Congruence and Welsh–English code-switching. *Bilingualism: Language and Cognition, 8*(03), 255-269.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knight (eds.) *Principles of frontal lobe function* (466-503). Oxford University Press: New York, NY.

Dijksta, T. (2005). Bilingual visual word recognition and lexical access. *Handbook of bilingualism psycholinguistic approaches, 54*, 179-201.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. *Handbook of bilingualism, 179*.

Dijkstra, T., & Van Heuven, W. J. (1998). The BIA model and bilingual word recognition. *Localist connectionist approaches to human cognition*, 189-225.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition, 5*(03), 175-197.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and language, 41*(4), 496-518.

Dijkstra, T., Van Heuven, W. J., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica 38*(3-4), 177-196.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition, 23*(2), 166-180.

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

Durston, S., Thomas, K. M., Yang, Y., Uluğ, A. M., Zimmerman, R. D., & Casey, B. J. (2002). A neural basis for the development of inhibitory control. *Developmental Science, 5(*4), F9-F16.

Dye, M. W., Green, C. S., & Bavelier, D. (2009a). Increasing speed of processing with action video games. *Current directions in psychological science, 18*(6), 321-326.

Dye, M. W., Green, C. S., & Bavelier, D. (2009b). The development of attention skills in action video game players. *Neuropsychologia, 47*(8), 1780-1789.

Ehri, L. C., & Ryan, E. B. (1980). Performance of bilinguals in a picture-word interference task. *Journal of Psycholinguistic Research, 9*(3), 285-302.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science, 270*(5234), 305.

Emmorey, K., Luk, G., Pyers, J., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science, 19*, 1201–1206

Engel de Abreu, P. M., 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

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

Esposito, A. G., Baker-Ward, L., & Mueller, S. T. (2013). Interference suppression vs. response inhibition: An explanation for the absence of a bilingual advantage in preschoolers' Stroop task performance. *Cognitive development, 28*(4), 354-363.

Fan, J., & Posner, M. (2004). Human attentional networks. *Psychiatrische Praxis, 31*(S 2), 210-214.

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *Neuroimage, 18*(1), 42-57.

Fan, J., Gu, X., Guise, K. G., Liu, X., Fossella, J., Wang, H., & Posner, M. I. (2009). Testing the behavioral interaction and integration of attentional networks. *Brain and Cognition, 70*(2), 209–220.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage, 26*(2), 471–479.

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*(3), 340–347.

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The Exposure Advantage Early Exposure to a Multilingual Environment Promotes Effective Communication. *Psychological science, 26*(7), 1090-1097.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, Massachusetts: MIT Press.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological science, 18*(10), 850-855.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and cognition, 9*(2), 288-307.

Ferrand, L., & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *Quarterly Journal of Experimental Psychology: Section A, 45*(3), 353-372.

Festman, J. (2012). Language control abilities of late bilinguals. *Bilingualism: Language and Cognition, 15*(03), 580-593.

Fias, W. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical cognition, 2*(1), 95-110.

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychological Research, 65*(4), 250-259.

Field, J. (2004). *Psycholinguistics: the Key Concepts*. United Kingdom, London: Routledge

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and speech, 24*(2), 125-146.

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain, 137*(10), 2811-2822.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Fodor, J. A. (2001). *The mind doesn't work that way: The scope and limits of computational psychology.* MIT press.

Folke, T., Ouzia, J., Bright, P., De Martino, B., & Filippi, R. (2016). A bilingual disadvantage in metacognitive processing. *Cognition, 150*, 119-132.

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 (1976). Accessing the mental *lexicon*. In F. Wales & E. Walker (Eds). *New approaches to language mechanisms* (p. 257-287). Amsterdam: North Holland.

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

Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet Neurology, 3*(6), 343-353.

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: models and data. *Trends in Cognitive Sciences, 8*(2), 87-93.

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science, 18*(10), 893-900.

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: a behavioral genetic analysis. *Developmental psychology, 47*(5), 1410.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*(2), 201.

Fritsch, T., Smyth, K. A., McClendon, M. J., Ogrocki, P. K., Santillan, C., Larsen, J. D., & Strauss, M. E. (2005). Associations between dementia/mild cognitive impairment and cognitive performance and activity levels in youth. *Journal of the American Geriatrics Society, 53*(7), 1191-1196.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology, 40*(2), 177.

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

Genesee, F., Boivin, I., & Nicoladis, E. (1996). Talking with strangers: A study of bilingual children's communicative competence. *Applied psycholinguistics,17*(04), 427-442.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 312–7 years old on a stroop-like day-night test. *Cognition, 53*(2), 129-153.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*(1), 13.

Gleason, J. B., & Bernstein, N. R. (1998). *Psycholinguistics.* Toronto: Harcourt Brace College Publishers.

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6*(01), 1-15.

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., & Acenas, L. A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 246.

Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: when more TOTs means better retrieval. *Journal of Experimental Psychology: General, 135(*3), 462.

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society, 13*(02), 197-208.

Gollan, T. H., Montoya, R. I., & Bonanni, M. P. (2005). Proper names get stuck on bilingual and monolingual speakers' tip of the tongue equally often. *Neuropsychology, 19*(3), 278.

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology, 16*(4), 562.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787-814.

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

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science, 22*(9), 1155-1164.

Goodenough F.L. (1926) *Measurements of Intelligence by Drawings*. New York: World Book Co

Goral, M., Campanelli, L., & Spiro, A. (2015). Language dominance and inhibition abilities in bilingual older adults. *Bilingualism: Language and Cognition, 18*(01), 79-89.

Graham, V. T. (1926). The intelligence of Italian and Jewish children in the habit clinics of the Massachusetts Division of Mental Hygiene. *The Journal of Abnormal and Social Psychology, 20*(4), 371.

Grainger, J. (1993). Visual word recognition in bilinguals. *The bilingual lexicon,6,* 11-26.

Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology, 39*(2), 295-319.

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: Elsevier Science Publishers B.V.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*(6939), 534-537.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and language, 27(*2), 210-223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition, 1(*02), 67-81.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515-530.

Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive development, 28*(1), 41-50.

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews, 55,* 88-97.

Grogan, A., Green, D. W., Ali, N., Crinion, J. T., & Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral Cortex, 19*(11), 2690-2698.

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual & Multicultural Development, 6(*6), 467-477.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and cognitive processes, 3*(3), 233-274.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. *Tutorials in bilingualism: Psycholinguistic perspectives,* 225-254.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and cognition, 1(*02), 131-149.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Harvard University Press.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism*. Basic Books.

Harley, T. A. (1990). Paragrammatisms: Syntactic disturbance or breakdown of control? *Cognition, 34*, 85-91.

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

Hartstra, E., Kühn, S., Verguts, T., & Brass, M. (2011). The implementation of verbal instructions: an fMRI study. *Human brain mapping, 32*(11), 1811-1824.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science, 15*(6), 409-414.

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

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language?. *Bilingualism: Language and Cognition. 1,* 213–229

Hernandez, A. E., & Reyes, I. (2002). Within-and between-language priming differ: Evidence from repetition of pictures in Spanish-English bilinguals.Journal of Experimental *Psychology: Learning, Memory, and Cognition, 28*(4), 726.

Hernandez, A. E., Bates, E. A., & Avila, L. X. (1996). Processing across the language boundary: a cross-modal priming study of Spanish-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(4), 846.

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.

Hill, T. D., Angel, J. L., Balistreri, K. S., & Herrera, A. P. (2012). Immigrant status and cognitive functioning in late-life: an examination of gender variations in the healthy immigrant effect. *Social science & medicine, 75*(12), 2076-2084.

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: exercise effects on brain and cognition. *Nature reviews neuroscience,9*(1), 58-65.

Holowka, S., Brosseau-Lapré, F., & Petitto, L. A. (2002). Semantic and conceptual knowledge underlying bilingual babies' first signs and words. *Language Learning, 52*(2), 205-262.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106(*1), 501-511.

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

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036.

Hulme, C., & Snowling, M. (1992). Deficits in output phonology: An explanation of reading failure? Cognitive Neuropsychology, 9(1), 47-72.

Ishigami, Y., & Klein, R. M. (2010). Repeated measurement of the components of attention using two versions of the Attention Network Test (ANT): stability, isolability, robustness, and reliability. *Journal of Neuroscience Methods, 190*(1), 117–128.

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

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A Test to Rapidly and Efficiently Assess the Spanish Vocabulary Size. *Psicologica: International Journal of Methodology and Experimental Psychology, 35*(1), 49-66.

Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and cognition, 4*(02), 169-178.

James, C. B. (1960). Bilingualism in Wales: An aspect of semantic organization. *Educational Research, 2*(2), 123-136. DOI:10.1080/0013188600020205

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7*(1), 2.

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press

Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. *Memory & Cognition, 15*(6), 531–543.

Jones, W. R. (1960). A critical study of bilingualism and non-verbal intelligence. *British journal of educational psychology, 30*(1), 71-77.

Jones, W. R., & Stewart, W. A. C. (1951). Bilingualism and verbal intelligence. *British Journal of Statistical Psychology, 4(*1), 3-8.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review, 17*(3), 213-233.

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*(5), 1243–1249.

Kandhadai, P., Hall, D. G., & Werker, J. F. (2016). Second label learning in bilingual and monolingual infants. Developmental science, 1-14.

Kapa, L. L., & Colombo, J. (2013). Attentional control in early and later bilingual children. *Cognitive Development, 28*(3), 233–246.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test (BNT).*Manual (2nd edition). Philadelphia: Lea and Fabiger.

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task switching training. *Developmental science, 12*(6), 978-990.

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

Kaushanskaya, M., & Marian, V. (2007). Bilingual language processing and interference in bilinguals: Evidence from eye tracking and picture naming. *Language Learning, 57*(1), 119-163.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*(4), 705-710.

Kennedy, S., McDonald, J. T., & Biddle, N. (2006*). The healthy immigrant effect and immigrant selection: evidence from four countries*. McMaster University.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological bulletin, 136*(5), 849.

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

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3-to 12-year-old Finnish children. *Developmental neuropsychology, 20*(1), 407-428.

Kousaie, S., & Phillips, N. A. (2012a). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials*. Brain research, 1446*, 71-90

Kousaie, S., & Phillips, N. A. (2012b). 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

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental science, 12*(1), 48-54.

Kovács, Á. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science, 325*(5940), 611-612.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and aging, 15*(1), 126.

Kray, J., Eber, J., & Lindenberger, U. (2004). Age differences in executive functioning across the lifespan: The role of verbalization in task preparation. *Acta Psychologica, 115*(2), 143-165.

Kreft, D., & Doblhammer, G. (2012). Contextual and individual determinants of health among Aussiedler and native Germans. *Health & place, 18*(5), 1046-1055.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language, 33*(2), 149.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. *Handbook of bilingualism: Psycholinguistic approaches,* 531-553.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition, 9*(02), 119-135.

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*(3), 416–430.

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. *The psychology of learning and motivation* 55, 229.

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*(03), 373-381.

Kroll, J.F., & de Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual: mapping form to meaning in two languages. In A.M.B. de Groot & J.F. Kroll (Eds.), *Tutorials in bilingualism* (pp. 169–199). Mahwah, NJ: Erlbaum.

Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science, 6*(3), 299–312.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology. General, 142*(2), 573–603.

Kuhn, R., Everett, B., & Silvey, R. (2011). The effects of children's migration on elderly kin's health: A counterfactual approach. *Demography, 48*(1), 183-209.

Laurent, A., & Martinot, C. (2010). Bilingualism and phonological awareness: the case of bilingual (French–Occitan) children. *Reading and Writing, 23*(3-4), 435-452.

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

Lee, M. W., & Williams, J. N. (2001). Lexical access in spoken word production by bilinguals: Evidence from the semantic competitor priming paradigm. *Bilingualism: Language and Cognition, 4*(03), 233-248.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(01), 1-38.

Levelt, W. J., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological review, 98*(1), 122.

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language the role of Retrieval-induced forgetting during Second-language acquisition. *Psychological Science, 18*(1), 29-34.

Lewis, D. G. (1959). Bilingualism and non-verbal intelligence: a further study of test results. *British Journal of Educational Psychology, 29*(1), 17-22.

Li, Y., Yang, J., Scherf, K. S., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and language, 127*(3), 452-462.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507-1515.

Livesey, D., Keen, J., Rouse, J., & White, F. (2006). The relationship between measures of executive function, motor performance and externalising behaviour in 5-and 6-year-old children. *Human Movement Science, 25*(1), 50-64.

Lobo, A., Ezquerra, J., Gómez, F., Sala, J. & 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.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*(5-6), 565-581.

Luk, G., De Sa, E. R. I. C., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition, 14*(04), 588-595.

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114*(1), 29-41.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition, 7*(6), 485-495.

Macizo, P., Bajo, T., & Martín, M. C. (2010). Inhibitory processes in bilingual language comprehension: Evidence from Spanish–English interlexical homographs. *Journal of Memory and Language, 63*(2), 232-244.

Mackie, M.-A., Van Dam, N. T., & Fan, J. (2013). Cognitive control and attentional functions. *Brain and Cognition, 82*(3), 301–312.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin, 109*(2), 163.

Macnamara, J. (1967). The bilingual's linguistic performance—a psychological overview. *Journal of social Issues, 23*(2), 58-77.

Macnamara, J., Krauthammer, M., & Bolgar, M. (1968). Language switching in bilinguals as a function of stimulus and response uncertainty. *Journal of Experimental Psychology, 78*(2), 208.

Magezi, D. A., Khateb, A., Mouthon, M., Spierer, L., & Annoni, J. M. (2012). Cognitive control of language production in bilinguals involves a partly independent process within the domain-general cognitive control network: Evidence from task-switching and electrical brain activity. *Brain and language,122*(1), 55-63.

Maguire, E. A., Spiers, H. J., Good, C. D., Hartley, T., Frackowiak, R. S., & Burgess, N. (2003). Navigation expertise and the human hippocampus: a structural brain imaging analysis. *Hippocampus, 13*(2), 250-259.

Mahon, M., & Crutchley, A. (2006). Performance of typically-developing school-age children with English as an additional language on the British Picture Vocabulary Scales II. *Child Language Teaching and Therapy, 22*(3), 333-351.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition, 6*(02), 97-115.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and language, 86*(1), 70-82.

Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin–English bilingual children. *Clinical linguistics & phonetics, 24*(4-5), 387-400.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 12 1- 157.

Marslen-Wilson, W. & Tyler, L. K. (1987). Against modularity. In J. L. Garfield (ed.), *Modularity in Knowledge Representation and Natural-Language Understanding*, Cambridge, MA: MIT Press

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word- recognition. *Cognition, 25*, 71-102.

Martin, C. D., Dering, B., Thomas, E. M., & Thierry, G. (2009). Brain potentials reveal semantic priming in both the 'active' and the 'non-attended' language of early bilinguals. *NeuroImage, 47*(1), 326-333.

Martin, C. D., Molnar, M., & Carreiras, M. (2016). The proactive bilingual brain: Using interlocutor identity to generate predictions for language processing. *Scientific reports, 6*.

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

Marzecová, A., Asanowicz, D., Kriva, L. U., & Wodniecka, Z. (2013). The effects of bilingualism on efficiency and lateralization of attentional networks. *Bilingualism: Language and Cognition, 16*(03), 608-623.

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology 2,* 22.

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: the role of backward inhibition. *Journal of Experimental Psychology: General,129*(1), 4.

McCarthy, D. (1946). Language development in children. In L Carmichael (Ed.), *Manual of child psychology*. New York: Wiley

McCauley, R. N. & Henrich, J. (2006). Susceptibility to the Müller-Lyer illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology, 19*: 79–101.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review, 88*(5), 375.

McNamara, J., & Kushnir, S. L. (1972). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior, 10*, 480–487.

Mead, M. (1927). Group intelligence tests and linguistic disability among Italian children. *School & Society, 1927* (25), 465-468.

Meaney, M. J. (2004). The nature of nurture: Maternal effects and chromatin remodeling. In J. T. Cacioppo, & G. G. Berntson (Eds.), *Essays in social neuroscience* (pp. 1–14). Cambridge: MIT Press

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: structural plasticity in the bilingual brain. *Nature, 431*(7010), 757-757.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of memory and language,40*(1), 25-40.

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

Michael, E., & Gollan, T.H. (2005). Being and becoming bilingual: Individual Differences and consequences for language production. In J.F. Kroll & A.M.B. de Groot (Eds.), *The handbook of bilingualism: Psycholinguistic approaches* (pp. 389-407). NY: Oxford University Press.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of abnormal psychology, 110*(1), 40.

Miller, N. A., & Kroll, J. F. (2002). Stroop effects in bilingual translation. *Memory & Cognition, 30*(4), 614-628.

Milne, B. J., Poulton, R., Caspi, A., & Moffitt, T. E. (2001). Brain drain or OE? Characteristics of young New Zealanders who leave. *The New Zealand medical journal, 114*(1141), 450-453.

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., ... & Shoda, Y. (2011). 'Willpower 'over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience, 6*(2), 252-256.

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

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Sears, M. R. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693-2698.

Molnar, M., Ibáñez-Molina, A., & Carreiras, M. (2015). Interlocutor identity affects language activation in bilinguals. *Journal of Memory and Language, 81,* 91-104.

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*(3), 308–330.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology, 25*(5), 531–546.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological science, 22*(11), 1425-1433.

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

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*(3), 359-377.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: general, 106*(3), 226.

Nelson, T.O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. Shimamura (Eds.): *Metacognition: Knowing about knowing* (pp. 1-25). Cambridge, MA: The MIT Press

Ng, E. (2011). The healthy immigrant effect and mortality rates. *Health Reports,22*(4), C1.

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

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York: Plenum

Normile, D. (2016). Nature from nurture. *Science, 351*(6276), 908-910.

Oller, D. K., & Eilers, R. E. (Eds.). (2002). *Language and literacy in bilingual children* (Vol. 2). Multilingual Matters.

Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. *Journal of child language, 24*(02), 407-425.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of educational psychology,98*(3), 554.

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., & Liu, Y. (2014). Conflict resolution in sentence processing is the same for bilinguals and monolinguals: The role of confirmation bias in testing for bilingual advantages. *Journal of Neurolinguistics, 27*(1), 50–74.

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

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

Palloni, A., & Arias, E. (2004). Paradox lost: explaining the Hispanic adult mortality advantage. *Demography, 41*(3), 385-415.

Paradis, M. (1980). The language switch in bilinguals: Psycholinguistic and neurolinguistic perspectives. *Languages in contact and conflict*, 501-506.

Paradis, M. (1981). Neurolinguistic organization of the bilingual's two languages. In J. E. Copeland & P. W. Davis (Eds.), *The Seventh LACUS Forum* (pp. 486–494). Columbia, SC: Hornbeam Press

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: general and applied, 76*(27), 1.

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language learning, 43*(1), 93-120.

Pelham, S. D., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals*. Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 313.

Penfield, W.; Roberts, L. (1959). *Speech and Brain Mechanisms.* Princeton: Princeton University Press

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current opinion in neurobiology, 15*(2), 202-206.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human brain mapping, 19(*3), 170-182.

Petitto, L. A., & Holowka, S. (2002). Evaluating attributions of delay and confusion in young bilinguals: Special insights from infants acquiring a signed and a spoken language. *Sign Language Studies, 3*(1), 4-33.

Petitto, L. A., & Kovelman, I. (2003). The bilingual paradox: How signing-speaking bilingual children help us to resolve it and teach us about the brain's mechanisms underlying all language acquisition. *Learning Languages, 8*(3), 5-19.

Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tétreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of child language, 28*(2), 453-496.

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-defined response sets. European Journal of *Cognitive Psychology, 19*(3), 395-416.

Pinker, S. (2003). *The blank slate: The modern denial of human nature*. New York: Penguin.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology, 22*(3), 415-422.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32*(1), 3–25.

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of experimental child psychology, 108*(3), 567-579.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied linguistics, 15*(1), 36-57.

Powledge, T. M. (2011). Behavioral epigenetics: how nurture shapes nature.BioScience, 61(8), 588-592.

Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical explorations, 9*(1), 29-43.

Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology,25*(5), 622-645.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and cognition, 13*(02), 253-262.

Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior research methods, 39*(4), 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*(1), 186–193.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge, MA: Blackwell.

Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language, 26*(4), 392-405.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review, 107*(3), 460.

Raven, J., Raven, J.C., & Court, J.H. (2004) *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment

Reynolds, A.G. (1991). The cognitive consequences of bilingualism. In A.G. Reynolds (Ed.), *Bilingualism, multiculturalism, and second language learning* (pp. 145–182). Hillsdale, NJ: Lawrence Erlbaum Associates

Ricketts, J., Nation, K., & Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235-257.

Ridley, M. (2003). *Nature via nurture: Genes, experience, and what makes us human*. HarperCollins Publishers.

Rigg, M. (1928). Some further data on the language handicap. Journal of Educational *Psychology, 19*(4), 252.

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology, 16*(4-6), 635-645.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1-3), 107-142.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(03), 465-485.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology, 7*(1), 17-24.

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

Rubio-Fernández, P., & Glucksberg, S. (2012). Reasoning about other people's beliefs: Bilinguals have an advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 211.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). *Development of attentional networks in childhood. Neuropsychologia, 42*(8), 1029–1040.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review, 89*(1), 60.

Saer, D. J. (1923). The effect of bilingualism on intelligence. *British Journal of Psychology. General Section, 14*(1), 25-38.

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International neuropsychological Society, 16*(5), 754.

Salthouse, T. A., & Mitchell, D. R. (1990). Effects of age and naturally occurring experience on spatial visualization performance. *Developmental Psychology, 26*(5), 845.

Salvatierra, J. L., & Rosselli, M. (2011). The effect of bilingualism and age on inhibitory control. *International Journal of Bilingualism, 15*, 26–37. doi:10.1177/1367006910371021

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

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition, 13*(02), 231-252.

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

Sanz, C. (2000). Bilingual education enhances third language acquisition: Evidence from Catalonia. *Applied psycholinguistics, 21*(01), 23-44.

Schade, U., & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production. II. The simulational perspective. *Journal of Psycholinguistic Research, 21*(6), 435-462.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language, 29*(1), 86-102.

Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology, 24*(5), 591-601.

Seidenberg, M. S., & McClelland, L. J. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*(4):527

Shallice, T., Burgess, P., & Robertson, I. (1996). The domain of supervisory processes and temporal organization of behaviour [and discussion].*Philosophical transactions of the Royal Society of London B: Biological sciences, 351*(1346), 1405-1412.

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia, 40*(6), 605-619.

Shin, H.B., & Kominski, R.A. (2010). *Language use in the United States*: 2007. Washington, DC: U.S. Department of Commerce Economics and Statistics Administration, U.S. Census Bureau.

Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition, 110*(1), 115-122.

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

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*(3), 214-241

Smith, F. (1923). Bilingualism and mental development. British Journal of Psychology. *General Section, 13*(3), 271-282.

Smith, M. C. (1997). How do bilinguals access lexical information. Tutorials in bilingualism: *Psycholinguistic perspectives*, 145-168.

Souchay, C., & Isingrini, M. (2004). Age related differences in metacognitive control: Role of executive functioning. *Brain and cognition, 56*(1), 89-99.

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within-group analysis approach. *Frontiers in Psychology, 2,* 183. http://dx.doi.org/10.3389/fpsyg.2011.00183.

Staff, R. T., Murray, A. D., Deary, I. J., & Whalley, L. J. (2004). What provides cerebral reserve? *Brain, 127*(5), 1191-1199.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 686.

Stins, J. F., Polderman, J. C., Boomsma, D. I., & de Geus, E. J. (2005). Response interference and working memory in 12-year-old children. *Child Neuropsychology, 11*(2), 191-201.

Stocco, A., & Prat, C. S. (2014). Bilingualism trains specific brain circuits involved in flexible rule selection and application. *Brain and language, 137,* 50-61.

Stocco, A., Lebiere, C., O'Reilly, R. C., & Anderson, J. R. (2012). Distinct contributions of the caudate nucleus, rostral prefrontal cortex, and parietal cortex to the execution of instructed tasks. *Cognitive, affective, & behavioral neuroscience, 12*(4), 611-628.

Strong, K., Tricket, P., Titulaer, I., & Bhatia, K. (1998). *Health in rural and remote Australia-the first report of the Australian Institute of Health*. Canberra: Commonwealth of Australia.

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

Sumiya, H., & Healy, A. F. (2004). Phonology in the bilingual Stroop effect. *Memory & cognition, 32*(5), 752-758.

Swain, M., Lapkin, S., Rowen, N., & Hart, D. (1990). The role of mother tongue literacy in third language learning. *Language, culture and curriculum, 3*(1), 65-81.

Swanson, H. L., Rosston, K., Gerber, M., & Solari, E. (2008). Influence of oral language and phonological awareness on children's bilingual reading. *Journal of School Psychology, 46*(4), 413-429.

Tao, L., Marzecova´, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology, 2*. doi:10. 3389/fpsyg.2011.00123

Team, R. C. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012.

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

Thomas, M. S., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language, 43*(1), 44-66.

Thomson, E. F., Nuru-Jeter, A., Richardson, D., Raza, F., & Minkler, M. (2013). The Hispanic paradox and older adults' disabilities: is there a healthy migrant effect? *International journal of environmental research and public health, 10*(5), 1786-1814.

Torrance, E. P., Gowan, J. C., Wu, J. J., & Aliotti, N. C. (1970). Creative functioning of monolingual and bilingual children in Singapore. *Journal of Educational Psychology, 61*(1), 72.

Tse, C. S., & Altarriba, J. (2014). The relationship between language proficiency and attentional control in Cantonese-English bilingual children: evidence from Simon, Simon switching, and working memory tasks. *Frontiers in Psychology, 5*, 1–18.

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. K. Peyton, D. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37–77). Washington, DC: Center for Applied Linguistics and Delta Systems.

van den Wildenberg, W. P., & van der Molen, M. W. (2004). Developmental trends in simple and selective inhibition of compatible and incompatible responses. *Journal of experimental child psychology, 87*(3), 201-220.

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

van Hell, J. G., & De Groot, A. M. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and cognition, 1*(03), 193-211.

van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*(4), 780-789

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of memory and language, 39*(3), 458-483.

Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition, 137*, 40-46.

Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Applied linguistics, 14*(4), 344-363.

Verhoeven, L. (2007). Early bilingualism, language transfer, and phonological awareness. *Applied Psycholinguistics, 28*(03), 425-439.

Vermeer, A. (1992). Exploring the second language learner lexicon. In J. H. A. L. De Jong & L. Verhoeven (Eds.), *The construct of language proficiency: Applications of psychological models to language assessment* (pp. 147–162). Amsterdam: John Benjamins.

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General, 145*(2), 246.

Wadsworth, M., Kuh, D., Richards, M., & Hardy, R. (2006). Cohort profile: the 1946 national birth cohort (MRC National Survey of Health and Development).*International journal of epidemiology, 35*(1), 49-54.

Wang, H., & Fan, J. (2007). Human attentional networks: a connectionist model. *Journal of Cognitive Neuroscience, 19*(10), 1678–1689.

Wang, S. L. (1926). A demonstration of the language difficulty involved in comparing racial groups by means of verbal intelligence tests. *Journal of Applied Psychology, 10*(1), 102-106.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392-393.

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.

Welsh, M. C. (2002). Developmental and clinical variations in executive functions. In D. L. Molfese & V. J. Molfese (Eds.), *Developmental variations in learning: Applications to social, executive function, language, and reading skills* (pp. 139–185). Mahwah: Lawrence Erlbaum.

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

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental psychology, 35*(1), 205.

Wilson, R. A. (2008). The drink you're having when you're not having a drink. *Mind & Language, 23*: 273–283

Wilson, R. S., Boyle, P. A., Yang, J., James, B. D., & Bennett, D. A. (2015). Early life instruction in foreign language and music and incidence of mild cognitive impairment. *Neuropsychology, 29*(2), 292e302.

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology, 79*(3), 1574-1578.

Woumans, E., Martin, C. D., Bulcke, C. V., Van Assche, E., Costa, A., Hartsuiker, R. J., & Duyck, W. (2015). Can faces prime a language? *Psychological science, 26*(9), 1343-1352.

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

Woumans, E., Surmont, J., Struys, E., & Duyck, W (in press). The longitudinal effect of bilingual immersion schooling on cognitive control and intelligence. *Language learning.*

Wu, Y. J., & Thierry, G. (2013). Fast Modulation of Executive Function by Language Context in Bilinguals. *The Journal of Neuroscience, 33*(33), 13533–13537.

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

Yin, X., Zhao, L., Xu, J., Evans, A. C., Fan, L., Ge, H., Tang, Y., Khundrakpam, B., Wang, J., & Liu, S. (2012). Anatomical substrates of the alerting, orienting and executive control components of attention: focus on the posterior parietal lobe. *PloS One, 7(*11), e50590.

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of abnormal psychology,118*(1), 117.

Yow, W. Q., & Markman, E. M. (2011). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development, 12*(1), 12-31.

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

Zhang, S., Morris, M. W., Cheng, C. Y., & Yap, A. J. (2013). Heritage-culture images disrupt immigrants' second-language processing through triggering first-language interference. *Proceedings of the National Academy of Sciences,110*(28), 11272-11277.

Laburpena euskaraz

Elebitasunaren ondorioak

Bi hizkuntzatan hitz egiteko gaitasunak egiten du pertsona bat elebidun, baina bi hizkuntzatan komunikatzeko ahalmena ez da elebiduna izatearen ondorio bakarra. Hamarkada batzuk badira, jada, elebidunen berezitasunek eta elebakarrekiko desberdintasunek zientzialarien interesa erakartzen hasi zirenetik. Desberdintasun hauek bi multzotan sailkatu ditzakegu. Hasteko, desberdintasun kuantitatibo bat dago: bakoitzak hitz egiten duen hizkuntza kopurua. Hizkuntza biak aztertuz eta elebakar batek egiten duen bere hizkuntzarekiko erabilerakin alderatuz, desberdintasun linguistikoak aurkitu ditzakegu. Bigarren desberdintasun motak ez dauka zer ikusirik kopuruarekin, hizkuntza bat baino gehiago jakiteak eskatzen duen goi-mailako prozesu kognitiboen inplikazioarekin baizik. Prozesu hauek, hizkuntzak doitasunez erabiltzeko behar den hizkuntza-kontrolean parte hartzen dutenak dira, baina ez dira hizkuntza-kontrolerako bakarrik erabiltzen. Hizkuntzaz bestelako egoeratan ere erabiltzen dira, eta agian elebitasunak eskatzen duen abilezia hoien erabilera finaren ondorioz, egoera orokorretan erabiltzen direnean ere findu egin daitezke, hobetuz. Beraz, bigarren talde honetan, desbertasun kualitatiboa izenda genezaken hontan, elebitasunaren ondorioz egoera orokorretako gaitasunetan sortu daitezkeen desberdintasun kognitiboak espera genitzake.

Orokorrean adostasuna dago elebitasunaren ondorio linguistikoen inguruan. Batzuk onuragarriak diren bitartean, beste batzuk eragozpenak uzten dituzte agerian. Lehenengo taldean, badirudi elebidunek bere hizkuntzekiko duten ezagutza txikiagoa dela elebakarrena baino (Bialystok, Craik, Green, & Gollan, 2009; Bialystok & Luk, 2012). Adibidez, elebidunek lexiko txikiagoa daukatela dirudi beraien hizkuntza bakoitzean elebakar batek baino (Verhallen & Schoonen, 1993; Vermeer, 1992; Perani et al., 2003; Portocarrero, Burright & Donovick, 2007), baita haurtzaroan ere (Mahon & Crutchley, 2006; Oller & Eilers, 2002, Bialystok, Luk, Peets, & Yang, 2010). Historikoki, honek garrantzia izan du, askotan ume batek dakien hitz kopuruaren ezagutza bere garapen linguistiko, kognitibo eta akademikoaren neurketa gisa hartu izan baita (Ouellette, 2006; Ricketts, Nation, & Bishop, 2007; Swanson, Rosston, Gerber, & Solari, 2008). Hortaz, oso desegokia izan daiteke ume elebidun batek dakizkien hitz kopuruan oinarrituta bere garapenarekiko ondorioak ateratzea, bereiziki bere bigarren hizkuntzan bada, ondorio okerrak atera baidaitezke. Honen inguruan, behin eta berriro frogatu da elebidunek eta elebakarrek ez dutela garapen kognitibo desberdinik (Baker & Jones, 1998; Cook, 1997; Hakuta, 1986).

Helduetan ere lexiko ezagutza desberdina aurkitu den arren (Bialystok, Craik, & Luk, 2008; Portocarrero, Burright & Donovick, 2007), askoz zailagoa da helduetan ezagutza hau zehazki neurtzea, kopuru hau faktore askoren menpekoa baita. Ondorioz, helduetan hipotesi hau sarrera lexikoan oinarrituz neurtzen da, hau da, hitzak sortzeko daukaten gaitasunean. Adibidez, irudi asko bata bestearen atzetik izendatuaraziz (Costa & Santesteban, 2004, Kaplan, Goodglass, & Weintraub, 1983). Bertan, elebidunek elebakarrek baino erantzun zuzen gutxiago ematen dituzte (Roberts, Garcia, Desrochers, & Hernandez, 2002; Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007), orokorrean mantsoago (Gollan, Montoya, Fennema-Notestine, & Morris, 2005) eta akats gehiagorekin (Bialystok et al., 2008). Sarrera lexikoa neurtzeko beste modu bat, kategoria konkretu bateko zenbat kandidatu esateko gai diren neurtzea da, denbora limite baten barruan. Adibidez, kategoria semantiko bat ("minutu bat daukazu ahal dituzun animalia izen guztiak esateko") edo fonologiko bat ("minutu batez, esan P hizkiaz hasten diren dakizkizun hitz guztiak"). Lehenengo kasua hitz kopuruaren neurketa garbiagotzat hartu ohi da, eta prozesu naturalago bat da, eguneroko bizitzan erlazio semantikoek bideratzen baidute hizketa. Kategoria fonologikoak zailagoak dira, eta prozesu desberdinen menpe daude (Grogan, Green, Ali, Crinion, & Price, 2009). Elebakarrek, normalki, elebidunek baina emaitza hobeak lortzen dituzte bi ariketetan, baina kategoria semantikoetan diferentziak haundiagoak dira (Bialystok et al., 2008, Gollan, Montoya, & Werner, 2002; Portocarrero et al., 2007).

Bestalde, garapenean hizkuntzak barneratzearen prozesuari begira, bi hizkuntza ikasten dituzten umeak eta bakarra ikasten dutenak konparatuz, elebitasunaren ulermena ere aldatzen joan da historikoki. Elebidunek garapen linguistiko motelago bat zutela defendatzen zen orain dela urte batzuk (Oller, Eilers, Urbano, & Cobo-Lewis, 1997; Pearson, Fernandez & Oller, 1993). Gainera, elebitasuna oso zaila zela uste zen, behintzat hizkuntza bakar batez jabetzea baino zailagoa (Macnamara, 1967; Torrance, Wu, Gowan, & Aliotti, 1970), eta hizkuntzen menperakuntzan atzerapenak edo nahasketak sortzearen arrazoia izan zitekela esaten zen. Azkeneko frogek diotenez, ez dago honelakorik. Elebidunek eta elebakarrek tankerako garapen linguistikoa dutela frogatu da, hizkuntzaren jabetzaren eta sorkuntzaren garapen prozesuko mugarrietara momentu berdinetan iristen direlarik (Petitto et al., 2001; Petitto & Kovelman, 2003; Petitto & Holowka, 2002), tankerako garapen semantikoa eta kontzeptuala erakutsiz (Holowka, Brosseau-Lapré, & Petitto, 2002), nahiz eta mugarri artean atzerapen nimio batzuk antzeman izan diren elebidunetan (Byers-Heinlein, Burns, & Werker, 2010).

Ikusten den bezela, elebitasunak dakartzan kalte linguistikoen ikuspuntua aldatzen joan da historikoki, eta gaur egun onartzen direnen eraigna, eguneroko bizitzan, oso txikia da, gehien bat kontestu esperimentaletara lotua baitaude.

Onura linguistikoen artean haundiena, ordea, elebiduna izateak mundua eta gure ingurua bi hizkuntzatan ulertzeko ahalmena dakarkigula da, eta bi hizkuntza hoietan gure burua ulertaraztea ere. Hortaz gain, elebitasunak abantailak dakartza hitz eta hizkuntza ikasketan ere, elebidunek elebakarrek baino erraztasun haundiagoz ikasten baidituzte hitz berriak euren hizkuntzetan (Kaushanskaya & Marian, 2009) eta hizkuntza berri batean (Cenoz & Valencia, 1994; Swain, Lapkin, Rowen, & Hart, 1990; Sanz, 2000). Hori gutxi balitz, atzerriko hizkuntza bat ikasterakoan, elebidunetan elebakarretan baino burmuin-sareen erabilera efizienteago bat aurkitzen da (Bradley, King & Hernandez, 2013).

Erraztasun honen jatorria, ziurrenik, haurtzarotik dator. Ume elebidunak malguagoak dira hitz berriak ikasterakoan, ume elebakarrek jarraitzen duten elkarrekiko esklusibotasunaren estrategia (hau da, objetu ezagun eta ezezagun bat ikustean izen ezezagun bat entzuten badute, objektu ezezagunaren izena dela ustea, Markman & Wachtel, 1998) ez baitute ume elebidunek aplikatzen (Kandhadai, Hall, & Werker, 2016). Hauek badakite objektu batek bi izen izan ditzakela, bat hizkuntza batean, eta bestea bestean. Historikoki ume elebidunek hizkuntzak bereizteko zailtasunak zituztela uste bazen ere, hori horrela ez dela frogatu da (Byers-Heinlein, Burns, & Werker, 2010; Bosch & Sebastián-Gallés, 2001). Ume elebidunek, soinu gabe grabatutako hiztunen bideoak ikusiz, umeen bi hizkuntzetako zeinetan hitz egiten ari diren bereizteko abilezia elebakarrek baina denbora luzeagoz mantentzen dute (Weikum et al., 2007), eta baita ezezagunak zaizkien bi hizkuntza badira ere (Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012).

Linguistikoki, onura asko daude eragozpen txiki batzuen frentean. Hala ere, ondorioak ez dira linguistikoak bakarrik, elebitasunaren eragina beste arlo kognitibotaraino iristen baita. Zientziak askotan frogatu du bizi esperientzia ugariren ondorioz gure jokaera eta abileziak aldatu daitezkela, bai neuralki eta baita kognitiboki ere. Bideojokoek ikuste-arreta hobetzen dute (Green & Bavelier, 2003; Feng, Spence, & Pratt, 2007), arkitektoek espazio-ikuste gaitasun hobeak dauzkate (Salthouse & Mitchell, 1990) eta Londoneko taxi gidariek burmuineko espazio-nabigazio atalak garatuagoak dauzkate (Maguire et al., 2000). Normalki heldutan ikasi eta egunean zehar gutxitan bizitzen ditugun esperientzia hauek burmuina eta jokaera aldatu badezakete, bi hizkuntza erabiltzeak eragin indartsu bat izan dezake, elebidun askok hizkuntzak txikitan ikasten baitituzte, eta egunero erabili. Elebitasunaren eragin kognitiboak ugariak dira ere, eta batzuetarako adostasuna lortu dugun bitartean, beste batzuetan oraindik eztabaida dago. Hemen ere, ondorio batzuk kaltegarriak diren bitartean, beste batzuek onura edo abantaila bat defendatzen dute (Bialystok, 2009).

Historikoki, elebitasunak eragin negatibo larriak zituela uste zen hainbat gaitasun kognitibotan, adibidez, adimenean (ikusi Darcy, 1946) "adimen nahasketa" (Saer, 1923), "pentsamendu zeheztasun gutxi" (Smith, 1923) edo "adimen atzerapena" (Goodenough, 1926) sortzen zuela pentsatzen zen, emaitza akademiko okerragoen kausa (Jones & Stewart, 1951). Handik urte gutxira, ikerketa horien akats ugari nabarmentzen hasi ziren (Jones, 1960; Darcy, 1953), elebidunen adimen gaitasuna elebakarrena baino baxuagoa zela defendatu zuten lan ugarietan (Graham, 1925; Lewis, 1959; Mead, 1927; Rigg, 1928; Wang, 1926) parte hartzaileen maila sozio-ekonomikoa ez baitzen kontuan hartu, edo elebidunek maila baxuago bat zuten (McCarthy, 1946). Honek garrantzi haundia du, adimen gaitasunean eragina izan dezakeen faktorea baita (James, 1960). Peal eta Lambert-ek (1962) elebidun eta elebakarrak faktore garrantzitsuetan berdindu ondoren (maila sozio-ekonomikoa, sexua, adina) elebidunek elebakarrek baina hobeto egin zutela argitaratu zuten. Honi "admimen malgotasuna" deitu zioten, eta aurkipen hau (eta geroago berdina erakutsi zutenak, adibidez Ben-Zeev, 1977) izan ziren "abaintaila elebidunaren" hipotesiaren aintzindari.

Kalte hauek zuzenak ez zirela frogatu ondoren, oso galera kognitibo gutxi aurkitu dira elebitasunari lotuta. Bakarrenetarikoa, elebidunetan aurkitu den metakognizio abilezia baxuagoa egongo litzake, bere buruaren jokaera elebakarrek elebidunek baina hobeto ebaluatzen dutela baitirudi (Folke et. al 2016).

Abaintailen inguruan ikerketa ugariagoa dago. Adibidez, ume elebidunak elebakarrak baina urtebete lehenago heltzen dira besteen usteak eta ikuspuntuak ondo ulertzera (adminaren teoria deitutakoa, Goetz, 2003; Kovács, 2009), uste egozentriko faltsuei aurre eginez (Rubio-Fernández & Glucksberg, 2012). Ikuspuntu kognitiboaz gain, ikuspuntu fisikoan ere ume elebidunak hobeak ageri dira, objektuak espazioan biratzeko abilezia hobea erakutsiz (Greenberg, Bellana, & Bialystok, 2013). Bi gauza hauen ondorioz, ume elebidunak hobeak dira ere instrukzioak azaltzen dituenaren ikuspuntu fisikotik ulertzen (Fan et al., 2015). Ziur aski, elebidunek inguruko pertsonen behar eta ezagutza linguistikoaz konsziente izatearen daukaten beharrak, besteen usteak eta ikuspuntuak hobeto ulertzen laguntzen dielako (Yow & Markman, 2011). Behar hauetara moldatzeko beharrak, egoera berrietara egokitzean ere elebakarrak baina hobeak egin ditu elebidunak, linguistikoak ez diren araudi berriak aplikatzerakoan elebidunak hobeto eta azkarrago moldatzen baitira (Stocco & Prat, 2014).

Abantaila elebidunaren hipotesia eta kritikak

Azkenik, elebitasunak funtzio exekutiboak hobetu ditzakelaren hipotesia, "abantaila elebiduna" bezala izendatua (Bialystok, 1999; Bialystok & Martin, 2004; Bialystok, 2011), eta bere inguruko debatea (Paap, Johnson & Sawi, 2015a; Duñabeitia & Carreiras, 2015) aurkitu dezakegu gaur egungo literaturan. Funtzio exekutiboen (FE) barne (Miyake & Friedman, 2012; Miyake et al., 2000) inhibizio-abilezia (erantzun indartsu bat erreprimitzeko gaitasuna), aldaketa-abilezia (lan eskema batetik bestera mugitzeko gaitasuna) eta berritze-abilezia (ingurua aztertu eta gure lan-memorian mantentzen dugun informazioa berritzen joateko gaitasuna) daude. Abilezia hauek oso garrantzitsuak dira gure egunerokotasunean (Mischel et al., 2011; Moffitt et al., 2011), gure autokontrola eta borondatea hoiengan oinarritzen baidira, erlazio sozialetan garrantzi handikoak (Friedman et al., 2007; 2011; Young et al, 2009). FE-n abilezia hauek entrenatuz hobetu daitezkenez (Moreno et al., 2011; Karbach & Kray, 2009; Kray & Lindenberger, 2000; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008), abantaila elebidunaren defendatzaileek elebitasuna FE-en entrenamendutzat hartu daitekela diote, elebitasunak FE-en erabilera bizi eta etengabeko bat eskatzen baitu, bi hizkuntzak ondo kontrolatzeko.

Kontrol honen beharra elebidun batek hitz egiten dituen bi hizkuntzak beti aktibo daudelako sortzen da (Dijkstra, Grainger & van Heuven, 1999; Sumiya & Healy, 2004;

Kaushanskaya & Marian, 2007; Thierry & Wu, 2007). Hori horrela izanik, hizkuntzaren control indartsu bat behar dute elebidunek, egoera bakoitzarako beharrezkoa den hizkuntza aukeratzeko eta komunikazio arrakastatsu bat lortzeko. Aukera ugari proposatu izan dira hizkuntza kontrol honen prozesuak azaltzeko (Penfield & Roberts, 1959; McNamara & Kushnir, 1971; Grosjean, 1988; 1997; 1998), baina modelu edo teoriarik onartuena gaur egun Green-ek (1998) proposatu zuen inhibizio kontrolaren (IK) teoría da. Bertan Green-ek hizkuntzen arteko lehiaz hitz egiten du (ikusi Abutalebi & Green, 2008, modelu honen moldakera neurokognitiboa ezagutzeko): bi hizkuntzek nahi dute azalera iritsi, aktibatu eta aukeratuak izan, erabiliak izateko. Beraz, inhibizioa aplikatu behar zaio nahi edo behar ez den hizkuntzari, erabiliko den hizkuntzaz sortuko den hizketan nahastu ez dadin. Modelu honek dionez, gure akzioen kontrola eta hizkuntzen kontrola mekanismo berdinen menpekoa da (Macnamara, Krauthammer, & Bolgar, 1968; Paradis, 1980; Norman & Shallice; 1986): atentzio sistema orokor batek eskema jakin batzuen (egin nahi den akzioaren eskema mentala, kasu hontan, hizkuntza eskema) erabilera eta aktibazioa gainbegiratzen du, helburu jakin batera iristeko. Helburua, hizkuntzen kasuan, informazioa komunikatzea da. IK modeluaren arabera, helburu honek kontzeptua sorrarazten du, eta hemendik sistema lexiko-semantikora bidaltzen da informazioa. Hizkuntza eskema bat aktibatzen denez, eskema horrek aukeratuko ditu sortu beharreko hitz (edo lemma) aproposak, erabili nahi ez diren hitzak inhibituz. Sisteman bi hizkuntza daudenean, egoera konplexuagoa da. Esaterako, transmititu nahi den kontzeptuak lemma kandidatuak aktibatzen dituenean, bi hizkuntzetakoak aktibatzen ditu (H1 eta H2). IK modeluak dioenez, atentzio sistemak erabili nahi ez den hizkuntza inhibitu egin behar du. Lemma ugariren aktibazioaren ondoren, batzuk asko inhibituko dira, eta inhibitu ez direnak aterako dira irabazle, sortzeko aukeratuak.

Beraz, IK modeluan oinarrituz, elebidunek jasan behar duten hizkuntza lehiari aurre egiteko inhibizioa asko erabili behar dute. Ondorioz, inhizioaren erabilera desberdina da elebakarretan eta elebidunetan: kontzeptu bat izendatu nahi denean, elebakarrek hitz helburu bakarra daukate, hizkuntza bereko gertukoak diren hautagaien artean aukeratu behar dutena, besteak inhibituz ("pescado" esan, eta ez "pez"). Elebidunek, ordea, hizkuntza arteko leiharekin ere tratatu behar dute ("pescado" esan, eta ez "pez" edo "arrain"), beste hizkuntza ere inhibituz. Abantaila elebidunak dionez, etengabeko hizkuntza-kontrol honen ondorioz elebidunak inhibitzaile hobeak bihurtuko lirateke (Bialystok, 2011). Horrela izanda, inhibizioa behar den egoera guztietan izango lirateke hobeak. Bizitzan, egoera askotan jokatu behar dugu atentzioz, garrantzitsuak ez diren estimuluak saihestuz eta garrantzitsuetan arreta jarriz: gidatzen goazela, ikasten gaudenean, erosketak egitean… une oro gabiltza gure atentzioaren erabilera eraginkor baten bila. Denari garrantzi berdina emango bagenioke, ez genukeelako jakingo garrantzitsua zer den bereizten. Beraz, galdera argi dago: hizkuntza-kontrolak eskatzen duen inhibizio lanaren ondorioz, hobeak al dira elebidunak inhibizioa behar den edozein egoera orokorretan?

Abantaila elebidunaren bila, psikologian aski ezagunak diren ariketa klasiko batzuk erabili izan ohi dira: flanker (Eriksen & Eriksen, 1974), Simon (Simon & Rudell, 1967) eta Stroop (Stroop, 1935) ariketak, adibidez. Ariketa guzti hauek inhibizio orokorrean oinarritzen dira, hizkuntza-kontrolaren kanpokoak. Guzti hauetan, estimulu kongruenteak (non aurkezten den informazio guztia bat datorren eskatzen den erantzunarekin) eta inkongruenteak (aurkezten den informazioaren zati batek eman behar den erantzunaren kontrakoa mesedetzen du) erabiltzen dira. Adibidez, Stroop ariketan parte hartzaileek pantailan ikusten dituzten hitzen kolorea esan behar dute ozenki, eta hitz hauen esanahia eta kolorea bat etorri daitezke ("gorria" hitza gorriz idatzia, kongruentea) edo ez ("gorria" hitza urdinez idatzia, inkongruentea). Simon ariketan parte hartzaileek lauki bat edo borobil bat ikusten dute pantailan, eta laukia ikustean "ezkerra" sakatu behar dutela, eta borobilarekin "eskuina", esaten zaie. Aurkezpen batzuk kongruenteak izan daitezke, laukia pantailaren ezkerrean aurkeztuz edo borobila eskuinean, hau da, sakatu behar den botoiaren alde berdinean. Beste aurkezpen batzutan inkongruentzia bilatzen da, laukia eskuinean aurkeztuz edo borobila ezkerrean. Tankerako logika jarraituz, flanker ariketak 5 gezi aurkezten ditu pantailan, denak lerro berdinean, bata bestearen ondoan. Parte hartzaileen eginbeharra erdian dagoen geziaren norantzakoa esatea da: ezkerreruntz edo eskuineruntz. Inguruko lau gezien arabera, baldintza kongruentea (denak zentzu berean, ← ← ← ← ←) edo inkongruentea (← ← → ← ←) sortu daiteke. Ariketa guzti hauetan, logika berdina da: inkongruenteak diren kasuak oso zailak dira, estimuluen parte garrantzitsu batek kontrako erantzuna mesedetzen duelako, eta askoz denbora gehiago behar dugu hauei erantzuteko. Kongruenteetan estimulu guztiek erantzun berdinaren alde egiten dutenez, oso errazak eta azkarrak egiten zaizkigu. Bi baldintza hauei erantzuna emateko behar dugun denbora eta akatsak konparatuz, baldintza inkongruenteak zenbat kaltetzen digun neurtu izan ohi da. Abantaila elebidunak dioenez, elebidunak hobeak dira balio ez dieten distrakzioak eta erantzun indartsuak inhibitzen (hizkuntzekin bezela), eta beraz baldintza inkongruenteen aurrean eragozpen gutxiago jasan beharko lukete. Honekin bat eginez, baldintza kongruente eta inkongruente arteko diferentzia txikiagoak aurkitu dira elebidunetan Stroop (Bialystok et al., 2008), Simon (Bialystok, 2006; Bialystok, Martin, & Viswanathan, 2005; Bialystok, Craik, Klein, & Viswanathan, 2004) eta flanker ariketetan (Costa, Hernández, & Sebastián-Gallés, 2008).

Emaitza hauek elebitasunaren ondoriotzat hartzen dira, elebidunen egunerokotasunean inhibizioa asko entrenatzearen ondorio zuzena, behar ez duten hizkuntzak ez eragozteko. Hala ere, hipotesi honen kontrako frogak indartzen ari dira. Adminari lotutako debatean bezela, gero eta ikerlari gehiagok diote abantaila elebidunaren jatorria ez dela elebitasuna, elebitasunari lotuta doazen beste faktore demografiko batzuk baino (Paap, Johnson, & Sawi, 2015a). Azken urteetako abantaila elebidunaren alde azaldu diren argitalapenak begiratuz, lan ugarik maila sozio-ekonomikoa kontuan hartzen ez dutela ikusten dugu (Bialystok & Martin, 2004) eta etorkinez osaturiko elebidunak bertako elebakarrekin konparatzen dituztela (Bialystok, Craik & Luk, 2008), kasu hoietan maila sozio-ekonomikoa ere kontuan hartu gabe (Bialystok & Shapero, 2005). Hori gutxi balitz, lurralde desberdinetako elebidunak eta elebakarrak ere aztertu eta konparatzen dira (Engel de Abreu, Cruz-Santos, Tourinho, Martin, & Bialystok, 2012). Alde batetik, badakigu maila sozio-ekonomiko altuago batek funtzio exekutibo (FE) hobeak dakartzala (e.g., Mezzacappa, 2004; Noble, Norman, & Farah, 2005), eta beraz faktore hori kontrolatu beharrekoa dela. Bestalde, kontuan izan beharrekoa da lurralde askok daukaten etorkin-politika zorrotzaren ondorioz, askotan atzerrira joatea lortzen dutenak adimen edo hezkuntza hobea dutenak izan ohi direla (Milne, Poulton, Caspi, & Moffitt, 2001; Wadsworth, Kuh, Richards, Hardy, 2006), FE abilezia hobeekin lotutako perfilak (Adelman et al., 2002; Arffa, 2007), alegia.

Izendatutako lan hoietan aurkitutako abaintala elebiduna hizkuntza perfilaren edo perfil horri lotutako beste faktoreen eraginez sortua izan zen jakiteko, faktore arrotz garrantzitsuetan berdinduak dauden elebakar eta elebidunak aztertu behar dira (Morton & Harper, 2007; Hilchey & Klein, 2011). Hortaz, aurkitutako abantaila elebitasunaren ondorioa bada, eta ez kontrolatu ez ziren faktoreena, abantailak agertzen jarraitu beharko luke behin taldeak berdinduta daudela ere. Azken emaitzek, ordea, ez dute abantailaren hipotesia bermatzen, faktore arrotzetan berdindutako elebakar eta elebidun taldeen artean ez baitugu atentzio-ariketetan inhibizio-abilezia alderik ikusten (Paap & Greenberg, 2013, Paap, Johnson, & Sawi, 2015a; 2015b).

Abantaila elebiduna aztertuz

Aurkezten dudan tesi hontan, hipotesi hau euskara eta gaztelania hitz egiten duten elebidun talde haundietan frogatu dut. Bi hizkuntzen erabilerak abantailak baldin badauzka, kontestu elebidun batean murgilduta, bizitza guztian zehar bi hizkuntza jakin eta erabili dituzten elebidunetan izan beharko litzake nabarien. Bizi garapenari begira, abantaila kognitibo hau errazago aurkitu izan da agure eta umeetan, heldu gazteetan baino (Bialystok, Martin, & Viswanathan, 2005), heldu gazteak beraien abilezien garapenaren maximoan daudelako (Hartshorne & Germine, 2015). Horren ondorioz, abantaila honek ezingo luke hoberenean dagoen abilezia bat gehiago hobetu. Beraz, abantaila bat nonbait aurkitu bagenezake, helduarotik urrun dauden taldeetan izan beharko litzake (Bialystok, Craik, & Luk, 2012).

Hasteko, abanataila elebiduna agureetan aztertu dut. Agureetan aurkitu izan den abantaila elebidunaren ebidentzien kritikak nabarmenak direnez parte-hartzaileen perfilen berdintzeari dagokionez (Bialystok, Craik, Klein, & Viswanathan, 2004; Gold, Kim, Johnson, Kryscio, & Smith, 2013), kritikak hobetzen saiatuz hipotesi bera lau esperimentuetan frogatu dut. Lehenengo bietan, Euskal Herriko 48 agure elebidun eta elebakar konparatu ditut ahozko eta zenbakizko Stroop ariketetan. Ahozko ariketa aurreko paragrafoetan azaldutako Stroop ariketa da. Zenbakizko bertsioan zenbaki bat azaltzen da pantailaren alde bakoitzean, biak tamaina desberdinekoak. Parte hartzaileek zenbaki haundiena (tamainan) zein zen esan behar zuten. Tamaina eta zenbakien balioa bat etorri liteke, baldintza kongruentea sortuz, edo ez, baldintza inkongruentean. Elebidunak eta elebakarrak aldagai garrantzitsuetan berdinduta zeuden (adina, adimena, hezkuntza...), eta beraz hizkuntz profila zen beraien arteko desberdintasun esanguratsu bakarra. Elebidunek ez zuten hobeto egin ariketa hauetan, ez baitzuten baldintza inkongruenteetan emaitza azkarrego edo hoberik eman. Hirugarren eta laugarren esperimentuetan, ariketa berdinak erabili nituen bigarren hizkuntza maila desberdina zeukaten 70 agure elebidun aztertzeko. Abantaila elebidunik baldin badago, eta hizkuntza kontrolean ona izateak FE-tan mesederik badauka, bigarren hizkuntzaren mailak garrantzia izan beharko luke. Gero eta hobea izan bi hizkuntzetan, hizkuntza kontrol hobea dagoela espero dezakegu, eta ondorioz, abantaila haundiagoa ariketa hauetan. Bigarren hizkuntzan zeukaten abilezia eta ariketan lortutako emaitzen artean ez nuen ordea inolako erlaziorik aurkitu. Beraz, zahartzaroko inhibizio abilezietan elebitasunak eragin positiborik ez dauakala dirudi, elebakarrek eta elebidunek berdin erantzuten baitute, ariketa hauetan, baldintza kongruente eta inkongruenteetara.

Batek esan dezake, ordea, agureen moteltze kognitiboak jota, abantailak baldin badaude ere zailagoak izan zitekela antzemateko. Horregatik, *continuum* demografikoko beste muturrera joan, eta abantaila elebitasuna harutzaroan ere aztertu dut. Zahartzaroan aurkitu ez izanaren arrazoia zahartzaroak dakarren moteltze kognitiboaren ondorioa baldin bada, abantaila umeetan garbia izan beharko litzateke, ez baitira helduak (non abileziak maximoan dauden) ezta agureak (moteltze prozesuan sartuta). Umeetan aurkitu izan den abantaila elebidunaren ebidentzietan ere (Martin-Rhee & Bialystok, 2008; Yang, Yang, & Lust, 2011) parte-hartzaileen berdintze arazoak egon direnez, hiru esperimentutan arazo hau konpontzen saiatuz hipotesi bera aztertu dut. Lehenengo bietan, agureekin erabili diren ariketa berdinak erabili dira, ahozko eta zenbakizko Stroop ariektak, zuhur berdindutako (adina, adimena, irakurmen gaitasuna…) 504 ume elebidun eta elebakarretan. Hirugarrenean, 360 elebidun eta elebakar ANT ariketan frogatu nituen, flanker ariketa oinarri bezala hartzen duen eta funtzio exekutiboak oso fin neurtu ditzaken ariketa bat. Berriz ere, ez zegoen alderik ume elebidun eta elebakarren artean hiru esperimentu hauetan, baldintza kongruente eta inkongruenteen aurrean berdin erantzuten zuten, FE-en garapen alderagarri bat erakutsiaz.

Bukatzeko, azkeneko talde demografikon aztertu dut abantaila elebidunaren hipotesia. Helduak dira orokorrean talde aztertuena zientzia kognitiboetan, eta hipotesi honekin berdin gertatzen da. Talde honekin erabili izan ohi dira ariketa mota gehien, eta abantaila aurkitu izan den arren (Bialystok, Craik & Luk; 2008; Bialystok & DePape, 2009) ez da beti horrela izan (Bialystok, Martin, & Viswanathan, 2005), lehen esan bezala, heldu hauen abilezia kognitiboak maximoan egotearen ondorioz, ziur aski. Hipotesi hau aztertzeko, 180 heldu elebidun eta elebakarrek umeek eta agureek egindako ariekta berdinetan frogatuak izan ziren (bi Stroop ariketak eta flanker), eta laugarren ariketa bat ere gehitu nuen, Simon ariketa, oso erabilia talde demografiko hontan. Berriz ere, elebakarrek eta elebidunek berdin jokatu zuten lau ariketatan, ez zen inon aurkitu erantzun eraginkorragorik baldintza inkongruenteetan, ez inolako abantailarik.

Emaitza hauek diotenez abantaila elebiduna (hau da, inhibizio abilezia hobeak edukitzea elebiduna izaeagatik), talde haundiak txukun antolatu eta faktore arrotz garrantzitsuetan berdindu ondoren, desagertu egiten da (Morton & Harper, 2007), eta hiptoesi berdina umeetan (Gathercole et al., 2014), helduetan (Paap & Greenberg, 2013) eta agureetan

(Kirk, Fiala, Scott-Brown, & Kempe, 2014; de Bruin, Bak & Della Sala, 2015) frogatu da munduko beste tokietan ere. Aurkikuntza hauen ondoren, teoriaren oinarriak berriro aztertu behar dira. Adibidez, abantaila elebidunaren oinarria den hipotesia berriz aztertu genezake: “hizkuntza-kontrolean erabili behar den inhibizioa eta hizkuntzarekin zer ikusika ez daukaten egoera orokorreko inhibizioa, inhibizio-abilezia berdina da” (Bialystok et al., 2005). Egia bada, agian elebitasunak gure gainean jartzen duen inhibizio beharraren karga ez da hain handia, eta eginarazten digun entrenamendu horrek ez gaitu maila gorenago batera bultzatzen. Azken finean, elebakarrek ere denbora guztian erabili behar dute inhibizioa, bai hizkuntzan (hitz egiteko forma aldatuz, familiakoei, lagunei, nagusiei, ezezagunei…) eta baita esparru orokorragoetan ere (gidatzen, lanean…), eta bi hizkuntzen kontrola inhibizio-abilezia orokorraren beste egin behar bat besterik ez da, beste askoren artean.

Honen guztiz kontrakoa den beste aukera bat badago, ordea, inhibizioan desberdintasunik aurkitzen ez dituzten emaitzak azaltzeko. Behar bada, hizkuntzaren kontrolean eta esparru orokorretan erabiltzen den inhibizioa ez da zertan berdina izan behar, eta abantailaren oinarria den hipotesia bera da oker dagoena. Ildo hontatik dijoazen lanak ez dira gutxi. Adibidez, pertsona berdinek hizkuntza-inhibizioa eta inhibizio-orokorra neurtzen duten ariketetan lortutako emaitzek ez dute korrelaziorik erakusten (Calabria, Hernández, Branzi, & Costa, 2012; Prior & Gollan, 2013; Calabria, Branzi, Marne, Hernández, & Costa, 2015; Branzi, Calabria, Boscarino, & Costa, 2016). Tesi honetan aztertutako partaideek inhibizio orokorreko ariketetan (ahozko eta zenbakizko Stroop ariketak, Simon ariketa eta flanker ariekta) lortutako emaitzak kontuan izanda, ariketen emaitzen arteko korrelazio analisiak egin, eta korrelazio esanguratsurik ez nuen aurkitu. Ez agureetan, ezta umeetan edo helduetan ere. Ariketa guzti hauek, nahiz eta denek egoera orkorreko inhibizioa neurtu, inhibizio mota desberdin baten menpe daudela ondorioztatu genezake, ez baitute beraien artean erlaizorik erakusten. Abilezia berdinaren menpe baleude, ariketa baten eta besteen emaitzak oso gertukoak izan beharko bailira. Hau horrela bada, hizkuntzarekin zerikusia daukaten eta ez daukaten ariketak inhibizio abilezia desberdinen menpe egongo lirake ere, guztiz bereiztuak. Burmuinari begira, badirudi hizkuntzari loturiko funtzio exekutiboak eta orokorragoak direnak ez direla zehazki atal berdin-berdinetan aurkitzen (Magezi, Khateb, Mouthon, Spierer, & Annoni, 2012; Branzi et al., 2015). Beraz, abilezia desberdinak badira, eta ez bata bestearen menpekoak edo orokorragoa den abilezia orokorrago baten menpekoak, elebitasunak hizkuntza kontrola hobetuko luke bakarrik, eta hori hobetzeak zerikusirik ez daukan inhibizio-abilezia konkretu bat hobetuko duela suposatzea jauzi haundiegia litzateke, berrikusi eta ziurrenik zuzendu beharrekoa. Honek abantaila elebidunaren hipotesia baliogabetuko luke.

Kontuan hartzekoa da, ordea, tesi honen asmoa ez dela elebitasunaren alde positiboak deuseztatezea edo txarra dela esatea. Elebitasunak ondorio onak eta txarrak izan ditzakela dirudi, baina lehenengoen garrantzia eta indarra askoz ere haundiagoa da, bigarrengoak kontestu esperimentaletan bakarrik aurkitu izan baitira, neurketa finetan isladatzen direnetakoak. Bi hizkuntza edo gehiago jakiteak ematen duen ikuspuntu irekieragatik, komunikazio erraztasunagatik eta esperimentu ugarietan aurkitu diren abantaila linguistiko eta kognitiboengatik ere, elebitasuna oso positiboa da, eta inhibizio abilezietan ez badago ere, abantaila elebiduna arlo askoretan oso esanguratsua da.