

Keynote Article

Cite this article: Leivada E, Westergaard M, Duñabeitia JA, Rothman J (2020). On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed? *Bilingualism: Language and Cognition* 1–14. <https://doi.org/10.1017/S1366728920000358>

Received: 10 January 2020

Revised: 12 March 2020

Accepted: 21 April 2020

Key words:

bilingualism; executive functions; neurocognition; Proteus phenomenon

Address for correspondence:

Evelina Leivada,

E-mail: evelina.leivada@urv.cat

On the phantom-like appearance of bilingualism effects on neurocognition: (How) should we proceed?

Evelina Leivada¹, Marit Westergaard^{2,3}, Jon Andoni Duñabeitia^{2,4} and Jason Rothman^{2,4}

¹Universitat Rovira i Virgili; ²UiT-The Arctic University of Norway; ³NTNU Norwegian University of Science and Technology and ⁴Universidad Nebrija

Abstract

Numerous studies have argued that bilingualism has effects on cognitive functions. Recently, in light of increasingly mixed empirical results, this claim has been challenged. One might ponder if there is enough evidence to justify a cessation to future research on the topic or, alternatively, how the field could proceed to better understand the phantom-like appearance of bilingual effects. Herein, we attempt to frame this appearance at the crossroads of several factors such as the heterogeneity of the term ‘bilingual’, sample size effects, task effects, and the complex dynamics between an early publication bias that favours positive results and the subsequent Proteus phenomenon. We conclude that any definitive claim on the topic is premature and that research must continue, albeit in a modified way. To this effect, we offer a path forward for future multi-lab work that should provide clearer answers to whether bilingualism has neurocognitive effects, and if so, under what conditions.

Introduction

Managing two linguistic systems in a single mind has been argued to leave its fingerprints on executive control (indirectly noted behaviourally) and foster neuroanatomical changes in the brain. Despite many studies claiming to show supportive evidence from sets of bilinguals tested across the lifespan (e.g., Bialystok, Craik & Luk, 2008; Bialystok, 2011; Luk, Bialystok, Craik & Grady, 2011; Lauchlan, Parisi & Fadda, 2013; Kroll & Bialystok, 2013; Costa & Sebastián-Gallés, 2014; Baum & Titone, 2014; Filippi, Morris, Richardson, Bright, Thomas, Karmiloff-Smith & Marian, 2015; Perani & Abutalebi, 2015; Burgaleta, Sanjuán, Ventura-Campos, Sebastian-Galles & Ávila, 2016; Blom, Boerma, Bosma, Cornips & Everaert, 2017; DeLuca, Rothman, Bialystok & Pliatsikas, 2019; DeLuca, Rothman, Bialystok & Pliatsikas, 2020), the nature and target of these bilingual effects are currently the subject of intense debate. Indeed, mixed reporting in the literature suggests that bilingualism does not (always) result in demonstrable differences in (cognitive) experimental performance (e.g., Morton & Harper, 2007; Paap & Greenberg, 2013; Paap, Johnson & Sawi 2015; Duñabeitia, Hernández, Antón, Macizo, Estévez, Fuentes & Carreiras, 2014; Antón, Duñabeitia, Estévez, Hernández, Castillo, Fuentes, Davidson & Carreiras, 2014; Ross & Melinger, 2016; Lehtonen, Soveri, Laine, Järvenpää, de Bruin & Antfolk, 2018). This is especially the case for commonly used tasks, such as the Flanker, Simon and Stroop, and with younger bilingual adults, a logical cohort for studies given the relative ease of access to them in university settings. Yet failure to find or replicate bilingual effects is not limited to these methods or populations. Thus, no one denies that bilingual effects, especially at the behavioural level, can have a PHANTOM-LIKE¹ QUALITY, or, as Costa, Hernández, Costa-Faidella and Sebastián-Gallés (2009: 135) put it, “now you see it, now you don’t”. The concern, then, becomes precisely how to reconcile this phantom-like appearance, interpreting what it tells us in general.

Studies in bilingualism follow more or less the same observational versus experimental divide found, for example, in the health sciences more generally (see Figure 1, adapted from Belluz & Hoffman, 2015). Unlike in the health sciences, however, where there is a clearer connection between the study types, there tends to be a more pronounced divide between observational and experimental studies in bilingualism; their use and (perceived) appropriateness go hand-in-hand with distinct questions related to diverse (yet complementary) paradigmatic approaches in linguistics, psychology, neuroscience and education.

While one can find both observational and experimental studies in cognitive neuroscience approaches to bilingualism, observational ones are relatively rare. Observational (cohort)

¹The term ‘phantom-like’ in no way implies that positive findings are (un)reliable. It merely points out the, as of yet, lack of determinacy in predicting *a priori* when effects might or might not obtain.

Strength of conclusions →

Experimental	SYSTEMATIC REVIEW & META-ANALYSIS	Collects all previous studies on the topic and systematically combines and evaluates their results.
	LONGITUDINAL CONTROLLED EXPERIMENTAL STUDY	Similar to one-time controlled experimental study that follows a cohort, testing them with the same experiments over time under the hypothesis that development has not yet reached a steady state and/or that various potential factors can change outcomes over time.
	ONE-TIME CONTROLLED EXPERIMENTAL STUDY	Offline and/or online experiments are administered to a group of bilinguals, typically compared against a baseline such as monolinguals and/or a cross-section of similar bilinguals that differ in a single main variable (e.g., age) or a set of quantified variables hypothesized to interact (e.g., age, proficiency and social background proxies).
Observational	LONGITUDINAL (COHORT) STUDY	Follows a group of bilingual people to track the role and weight of variables to outcomes over time, either via questionnaires and/or natural (non-elicited) production of language.
	CASE-CONTROL STUDY	Compares histories of a group of bilingual people for any number of factors/variables to another group of people, often monolingual speakers of one of the languages of the bilinguals.
	CROSS-SECTIONAL SURVEY	Assesses the prevalence of a variable or an outcome in a broad population of bilinguals at a single point in time, either via questionnaires and/or natural (non-elicited) production of language.
	CASE REPORTS	Detailed histories of individual cases of bilinguals, examining any number of issues related to bilingual acquisition and/or related sociological issues.

Fig. 1. Study Type Hierarchy related to Strength of Conclusions

studies have been significant in the literature examining potential links between bilingualism and neurodegeneration, for example, studies correlating later Alzheimer's/dementia diagnosis with bilingualism (e.g., Bialystok, Craik & Freedman 2007; Craik, Bialystok & Freedman, 2010; Chertkow, Whitehead, Phillips, Wolfson, Atherton & Bergman, 2010; Alladi, Bak, Duggirala, Surampudi, Shailaja, Shukla, Chaudhuri & Kaul, 2013; Yeung, St. John, Menec & Tyas, 2014; Lawton, Gasquoine & Weimer, 2015). Nevertheless, the overwhelming majority of studies dealing with bilingualism and neurocognition are experimental, typically of the one-time controlled type (cf. Figure 1). Although there are some discrepant conclusions across studies, the crux of the evidence for the phantom-like appearance of bilingual effects comes from the experimental literature related to executive functions. It is not only the case that there are studies showing bilingual effects and studies that fail to replicate findings, some recent meta-analyses also suggest that there is serious reason to be skeptical of any deterministic bilingual effects on cognition. The bird's eye view that meta-analyses/systematic reviews offer has led

several scholars to the conclusion that a generalized bilingual effect is exaggerated in frequency and is more likely a byproduct of a confirmation bias in general and/or a bias towards not publishing null results (e.g., Paap et al., 2015; Lehtonen et al., 2018). In fact, Lehtonen et al.'s (2018) analysis claims that when relevant unpublished data are included and a number of study, task, and individual participant related variables are properly considered, bilingual effects on inhibition, shifting and working memory disappear after correcting estimates for publication bias.

Given the weight that systematic reviews and meta-analyses have in the hierarchy of STRENGTH OF CONCLUSIONS as schematized in Figure 1, they should be in a privileged position to offer significant insights. Nevertheless, it is not the case that all systematic reviews and meta-analyses reach the same conclusions, a quandary that might relate to the current debates regarding the appropriateness of some approaches to synthesis studies and meta-analyses (see Ioannidis, 2016; Papatheodorou, 2019). Hilchey & Klein's (2011) meta-analysis of bilingual data from interference tasks, for example, showed no greater performance

in bilinguals. However, they demonstrated that bilinguals were generally better in both compatible and incompatible trials to the same magnitude. Thus, while they did not conclude that data support a bilingual effect on interference resolution per se, as claimed in many individual studies, they pointed out that the combined results “suggest bilinguals do enjoy a more widespread cognitive advantage (a bilingual executive processing advantage) that is likely observable on a variety of cognitive assessment tools but that, somewhat ironically, is most often not apparent on traditional assays of non-linguistic inhibitory control processes” (Hilchey & Klein, 2011: 625). In a similar vein, van den Noort, Vermeire, Bosch, Staudte, Krajenbrink, Jaswetz, Struys, Yeo, Barisch, Perriard, Lee and Lim’s (2019) review of 46 original studies on bilingualism and cognitive control also found a spread of results (54.3% beneficial effects, 28.3% null effects and 17.4% evidence against bilingual effects). Their analysis showed that issues of compatibility across studies, often methodological (participant selection, tasks used, individual differences not considered, lack of longitudinal designs), had good explanatory power for cross-study disparities. While they claimed to find some evidence overall for bilingual effects, they highlight that a serious risk for (unintentional) biases exists in both a confirmation and a disconfirmation direction.

On the whole, recent meta-analyses and systematic reviews give cause for reflection, if not concern. While we have no doubt that individual studies have been done to high standards, what can be concluded from bringing them together is not at all clear. Of course, not all meta-analyses and systematic reviews are created equal. That which can be understood (better) from a meta-analysis or systematic review is inherently related to the actual appropriateness of bringing included data sets together in the first place. Data must be similar enough to warrant their being combined. Determining what similar enough means is of no small consequence. Failure to get this crucial condition right could translate to comparisons of proverbial apples to oranges, the blending of which fails in the most essential ways to ensure confidence for meaningful conclusions that sound meta-analyses should provide. In light of the provisos discussed in van den Noort et al.’s (2019) work, if methodological differences reduce the similarity/comparability of data sets to a significant degree, then we must consider what consequences these have for meta-analyses and systematic reviews. Furthermore, since bilingualism itself is defined distinctly in many studies, i.e., often not treated as the spectrum it is, we must ponder what the consequences are of collapsing data across studies with participants of vastly different bilingual profiles.

In light of the above, how do we move forward in the general program of trying to determine what, if any, effects bilingualism has on the mind and brain? The stakes are high because determining when evidence has reached a critically sufficient mass to abandon an established trend almost always has manifold implications. Given the potential benefits for individual health and society that bilingual effects on neurocognition could entail, we must be absolutely positive that there are no effects before determining it is time to abandon the search. At the same time, it is worth pondering whether the presence of suggestive findings that are not consistently replicated across labs fully supports the admittedly strong arguments put forward in relation to a seemingly causal relationship between bilingual experience and neuroprotection. The real question is: do we truly know enough yet to definitively claim that positive findings of bilingual effects on neurocognition are nothing more than an artefact of methodology

and confirmation bias? If the answer is an unequivocal ‘yes’, it is time to abandon the endeavour altogether. If the answer is ‘no’ or if we are simply ‘unsure’, then the only responsible conclusion is to continue. However, we cannot afford to continue blindly: some basic common rules should be agreed upon by researchers in the field. The intrinsic value of asking the question in the first place is the opportunity it provides for consolidating what we know or have learned between intervals of taking stock, to be able to move forward with increased wisdom, humility and precision. If the general program investigating the possibility of bilingual effects is to continue, as we will make a case for in the remainder of this paper, it must adapt to avoid circularity, finding a good balance between revolution and evolution in the findings. We need to establish and agree on a common ground through which labs across the world work in complement to collectively narrow in on a better understanding of the common goal: determining the conditions under which, if any, bilingualism has an effect on the mind and brain. This is not a trivial endeavour. Such a push cannot be circumvented by big data alone, unnuanced in considering the dynamic nature of the bilingual experience and its potential determinism, as in Nichols, Wild, Stojanoski, Battista and Owen (2020). Power in our work is of crucial importance. However, power cannot take precedence over nuance, especially when neither need to be sacrificed, as we discuss in detail below, offering suggestions on how to achieve this. Alternatively, big data runs the risk of adding to, rather than working towards resolving, the relevant debates.

The present article is an attempt at carving out a path to do just that. Without pretence or pretext, we, a team of scholars with distinct inclinations about how the cards will fall in the end, join forces to unpack key issues related to the present debate. While we do not completely agree on how to view and interpret all available data, we offer facts for consideration as neutrally as possible. We critically discuss a subset of factors that might contribute to the phantom-like appearance of bilingual effects, the consequence of which requires a reshaping and reconsideration of how we approach our object of study and any conclusions that have been made about it to date: (i) the heterogeneity of the term ‘bilingual’, (ii) sample size effects and variability in power, (iii) task effects and (iv) the complex dynamics between an early publication bias that favours positive results and the subsequent Proteus phenomenon. We are united in our desire to outline a tangible way forward for better standards and cross-lab collaborations capable of yielding maximally comparable and reliable data.

Setting the context: Initial thoughts on the phantom-like appearance of bilingual effects

Phantom-like appearances of effects are not unique to the domain of bilingualism and cognition. In fact, virtually all areas of academic inquiry that have moved beyond initial findings suggestive of a robust effect produce studies offering positive, null and even negative results, increasingly so as researchers test the limits of the initial findings (e.g., de Bruin & Della Sala, 2015). As concerns bilingualism and cognition, the present debate is not (or should not be) about the existence of bilingual effects in general, under constrained conditions only or no generalizable effects at all, but rather what we should responsibly conclude from the totality of conflicting data.

As always, terminology matters. In the present case, in our view, the imprecision of a particular descriptive term attributed to apparent bilingual effects significantly contributes to misunderstanding and miscommunication. The term ‘BILINGUAL ADVANTAGE’

is omnipresent in the literature, yet entirely inaccurate even if it were to refer to a bona fide and generalizable bilingual effect on neurocognition. A recent search in Scopus® at the time of writing this article showed that there are currently more than 300 research articles including the term 'BILINGUAL ADVANTAGE' either in the title, the keywords or the abstract. Moreover, instead of diminishing the literal reference to that term in light of the recent debate, during the year 2019 the specific mentions to 'bilingual advantage' have increased by nearly 30%. Claiming an effect or anything as an advantage is often *a priori* spurious because its qualification as such depends largely on specific perspectives and interpretations of (in our case, behavioral) corollaries themselves. There is likely a trade-off to accommodating adaptations on the mind and brain induced by intense and prolonged experiences. What many or most would view, in isolation, as advantageous in one cognitive domain can come at a cost to another. Conversely, what might seem to have real advantages in practical terms at present, could be viewed completely oppositely down the line as (external) contexts change.

Let us consider a tangible example. If under certain conditions bilingualism contributes to both cognitive and neural reserves that translate into protection against or compensation for typical or pathological cognitive ageing, understanding this as an advantage would at best be context-dependant and temporal. Helpful as it might be, the observation that bilingualism correlates with delayed emergence of symptoms of Alzheimer's/dementia and, thus, later diagnosis by 4–6 years compared to monolinguals is objectively not an advantage *per se*. Despite media headlines, no one has ever claimed that life-long bilingualism somehow cures or prevents Alzheimer's/dementia. Rather, hypothesized to result from the bilingualism-induced accruing of the abovementioned reserves, neurodegeneration is compensated for in behaviour, without stopping or reversing underlying progression in the brain. Such diseases are marked by a preclinical phase where the pathology exists and is traceable based on specific biomarkers, even in cognitively normal individuals with complete asymptomatic behavior (e.g., for Alzheimer's see Aisen, Cummings, Jack, Morris, Sperling, Frölich, Jones, Dowsett, Matthews, Raskin, Scheltens & Dubois, 2017; Preische, Schultz, Apel, et al., 2019). And so, bilinguals, on average, show later onset of overt symptoms – but not underlying neuritic plaquing *per se* – relative to monolinguals and thus, diagnosis is set back. At present, with few available treatments, this means longer quality of life and is logically viewed as advantageous. However, in the future, later overt signs of behavioural symptoms might prove problematic. All things being equal, nothing would need to change for this so-called advantageous happenstance to turn rather disadvantageous; delayed symptoms translating to later diagnosis could derail interventions when such become available.

In any case, as scientists we do not (or should not) engage with reductionist terms to complex and dynamic entities. They not only oversimplify matters at hand, but contribute in no small part to the creation of contexts, especially in the absence of reliable replication, for polarization in all possible directions. For this reason, although the term is often used in the literature we discuss, we will not use 'BILINGUAL ADVANTAGE' in the remainder of this paper. In fact, we strongly recommend its disuse in favor of more neutral terms. Herein, we use the term 'bilingual effects' to refer to the impact bilingualism may have on neurocognition.

How can dichotomous conclusions – and many intermediary ones – about the very existence of a bilingual effect on neurocognition be argued in light of the same data available to all? Just as

an affirmative position has the clear burden of accounting for why there is a phantom-like appearance of the bilingual effect on cognition, a negating position has an equal burden of explanation for the many studies that do find behavioral evidence in support. Evidence of absence in some, even many, studies should not necessarily be understood as absence of evidence overall. It thus seems that any generalized conclusion, in the positive or negative, is at present precipitous. Hinging conclusions for this important question on the basis of commonly used executive function tasks, most typically with participants at peak levels of cognition in young adulthood, is not the best adjudication (e.g., Bialystok 2016, 2017). Given issues related to potential task-granularity effects in populations of peak-level cognition (young adults), it is interesting to consider the literature on neuro-anatomical adaptation that runs in parallel to the executive function literature.

If the mental juggling inherent to bilingualism affords cognitive and neural reserve, it is reasonable, given that adult brains remain highly plastic (see Fuchs & Flugge, 2014 for review), to expect measurable physiological changes to the brain. Due to the nature of neuro-imaging, which essentially provides a snapshot of structure and functional connectivity of the brain, we might expect more consistent results in this field. Given the claimed underlying mechanisms at play coupled with topographical roadmaps from the language processing and cognitive neuroscience literatures, one can make precise predictions that can be reasonably linked to bilingual experiences (see Pliatsikas, 2019a for review). According to Paap et al. (2015: 265),

brain imaging studies have made only a modest contribution to evaluating the bilingual-advantage hypothesis, principally because the neural differences do not align with the behavioral differences and also because the neural measures are often ambiguous with respect to whether greater magnitudes should cause increases or decreases in performance.

Paap et al. (2015) rightly point out that neuro-anatomical differences do not always align with behavioral performances. However, one should not expect that it would for several reasons, not the least given issues of granularity with executive function tasks themselves and the fact that positive effects of bilingualism could result in both expansion (evidence of greater involvement) and reduction (evidence of increased efficiency over time) of cerebral areas/neurological pathways (see Pliatsikas, 2019a; b for discussion). Indeed, monolinguals and bilinguals might perform the same behaviorally, but neuroimaging evidence can reveal if the relative effort for both groups is equal or if one group exerts less effort for the same performance. The goal of a good portion of neuro-imaging studies, for example all resting state ones, is not to examine correlations between neuro-anatomical change and task performance. Rather, they stand in complement to investigate the extent to which brain regions implicated specifically in language processing and relevant executive functions are affected. For fMRI studies with executive function tasks, it is true that changes can be noted without specific effects in performance, but again the aim of such studies is not predicated on an expectation for behavioral performance correlations. The goal, rather, is to test if recruitment in neuronal pathways in predictable areas of the brain is differentially affected and can be related to increased efficiency, whether or not behavior correlates. Very recent neuro-imaging studies, in fact, provide good evidence for the aforementioned and show how specific experiences related to bilingualism (exposure, domains of use, etc.) correlate to greater probability at the individual level of neuro-anatomical change/

more efficient neuronal recruitment during behavioral task performance (see Dash, Berroir, Joannette & Ansaldo, 2019; DeLuca et al., 2020; Sulpizio, Del Maschio, Del Mauro, Fedeli & Abutalebi, 2020a).

Indeed, a growing number of studies in recent years attest to adaptations in bilingual brain network activity and structure, crucially in areas implicated in language control and processing commensurate with bilingual language use (see Pliatsikas, 2019b for review). Language and executive control/processing are served by overlapping neural regions and networks (De Baene, Duyck, Brass & Carreiras, 2015; Green & Abutalebi, 2013), and demands on the language control system have been found to affect domain-general control (Parker Jones, Green, Grogan, Pliatsikas, Filippopolitis, Ali, Lee, Ramsden, Gazarian, Prejawa, Seghier & Price, 2012). Yet the relationship between brain structure and cognitive function is far from being clear, and so is the mechanistic explanatory power of structural neuroimaging studies per se (see Duñabeitia & Carreiras, 2015). As discussed immediately above, differences in patterns of neural recruitment are not consistently found to translate to differences in task performance, and inconsistencies exist between studies with respect to where and how bilingualism affects neural recruitment in cognitive control processes (Luk, Anderson, Craik, Grady & Bialystok, 2010; Costumero, Rodríguez-Pujadas, Fuentes-Claramonte & Ávila, 2015; García-Pentón, Fernández García, Costello, Duñabeitia & Carreiras, 2016; Pliatsikas & Luk, 2016). Nevertheless, neuroanatomical adaptations are reliably shown in studies examining bilinguals of all ages, even the illusive young adult age range at peak levels of cognitive performance. Neuro-anatomical imaging with (structural) MRI is not subject to task performance effects in the way that executive function tasks are. And so, the relative consistency of findings examining brain adaptations directly suggests that bilingualism, at least under conditions of active use and engagement (Luk & Bialystok, 2013; Li, Legault & Litcofsky, 2014; DeLuca et al., 2019), has effects consistent with claims that it leaves an indelible mark. While it could be the case that there is no reliable effect of bilingualism on executive functions, we need to reconcile the phantom-like appearance in the behavioral domain with the neuro-anatomical literature, to the extent that the implied underlying mechanisms are one and the same. This need does not pertain only to bilingualism research. It is a larger issue of structure-behavior relationships more generally; according to recent research suggesting that finding consistent and significant associations between behavioral performance and brain morphology is unlikely (Masouleh, Eickhoff, Hoffstaedter & Genon, 2019).

Notwithstanding the above, if we are to move forward in this general program, we must understand better what variables drive and lead to bilingual mind/brain adaptations, thus differentiating sets of individuals and groups from one another. Several factors have been identified as positively related to the conferment of bilingual effects, for example, (i) level of education, (ii) degree of language proficiency, (iii) age of onset of bilingualism, and (iv) frequency of use of the two languages (Guzmán-Vélez & Tranel, 2015 *inter alia*). This list is not exhaustive, and one of the goals of the present work is to discuss another set of factors that, coupled with others, may help us to understand better the phantom-like appearance of bilingual effects in the literature.

Importantly, ALL these factors offer a PROBABILISTIC perspective into the occurrence of mind/brain adaptations, as attested through different tasks and in different language communities, not a DETERMINISTIC one. A possibility that has not received

sufficient attention so far is that different occurrences/degrees of bilingual effects could be the outcome of a DISTINCT INTERACTION OF FACTORS, rather than boil down to the same (sub)set of deterministic and universally reliable variables. This is not to say that these factors cannot be universally or reliably related to bilingual effects. The claim is that in a multi-causal world situation, the operation of complex, multivariate patterns is the norm, and factors of influence often push in opposite directions (Liebersohn, 1991). In the present case, this entails that across different (i) conditions of testing, (ii) populations, and (iii) cognitive measures, the influence of a cluster of factors such as high level of education and/or high degree of language proficiency in two languages² can be outweighed by another cluster of factors such as type of bilingual trajectory, incidence, and context of language use (Luk & Bialystok, 2013; Kroll & Chiarello, 2016; Li et al., 2014; Bak, 2016a; Bialystok, 2016; Gullifer, Chai, Whitford, Pivneva, Baum, Klein & Titone, 2018; DeLuca et al., 2019; 2020; Beatty-Martínez, Navarro-Torres, Dussias, Bajo, Guzzardo Tamargo & Kroll, 2019). If some of these factors eventually cancel each other out or were never available in proportions sufficient to trigger neurocognitive adaptations, it would follow that different studies on bilingual cognition could reach contradictory results because of sampling issues, even when they employ the same tasks or recruit their subjects from the same linguistic community.

One must also contemplate the possibility that the phantom-like appearance of the bilingualism-induced behavioral effects relates to factors that are not strictly related to bilingualism. A number of leisure or social activities can lead to enhanced cognitive performance, e.g., music training (Bialystok & DePape, 2009; Linnavalli, Putkinen, Lipsanen, Huotilainen & Tervaniemi, 2018). We agree with Valian (2015) that potential cognitive effects of bilingualism COMPETE with other sources of adaptation in both monolingual and bilingual populations, and in the event that the other sources are sufficiently plentiful, bilingual effects may either be nullified or capturing them with traditional executive function tasks or neuroimaging might be compromised. For example, a well-known set of seminal studies by Maguire and colleagues (e.g., Maguire, Burgess, Donnett, Frackowiak, Frith & O'Keefe, 1998; Maguire, Gadian, Johnsrude, Good, Ashburner, Frackowiak & Frith, 2000) have shown similar neuroanatomical adaptations for taxi driver brains – specifically in the hippocampus – presumably because the skills needed to navigate involve some of the same systems that bilingualism is argued to engage. It could be the case that a ceiling effect would be reached such that monolingual and bilingual taxi cab drivers would show no or negligible differences; bilingualism would potentially confer no more changes to the mind/brain in this case because the activities involved in constant and expert navigation already max out potential effects. This is not limited to taxi cab drivers, of course; all activities that engage the same systems that subsume executive functions may provide similar opportunity. The people who are truly experts in these many activities could also reach ceiling effects, obscuring the role that bilingualism may have otherwise had. As we have no way to know if any given sample contains more or less of such people, this ceiling effect could give rise to some of the phantom-like results documented in the literature. And put differently, if bilingualism is a form of maximal language expertise, then the

²But see DeLuca, Rothman and Pliatsikas (2018) for discussion of why proficiency after a minimal threshold might lose its predictive validity, depending on what the underlying mechanisms involved in bilingual effects to the mind/brain turn out to be.

obscuring of the effects could take place in the opposite direction too. All in all, expertise in a given domain is often at the core of outstanding effects in certain cognitive skills or brain structural properties, be it of mathematical (e.g., Jeon, Kuhl & Friederici, 2019), musical (e.g., Saari, Burunat, Brattico & Toiviainen, 2018), or any other nature, including linguistic, and we are far from understanding the manner in which different forms of expertise conspire to shape the brain and neurocognitive processes (see Debarnot, Sperduti, Di Rienzo & Guillot, 2014).

Having established the general picture of the behavioural and neuroanatomical issues that surround the adaptations and effects bilingualism may induce on neurocognition, we are left with a few remaining aims. The first is to examine some examples of potentially confounding methodological factors. The second is to provide a concrete path for moving forward, keeping in mind the provisos that obtain in the course of undertaking the first aim.

The heterogeneity of the term ‘bilingual’ and its implications for meta-analyses

The term ‘bilingual’ is an umbrella construct that can host quite different populations. Consider for example the following extreme definitions:

- (1) Any person who knows at least a few words in a language other than the maternal variety is bilingual (Edwards, 2004: 7)
- (2) Bilingual is a person that has native-like control of two varieties (Bloomfield, 1933: 56)

There are many ways of being bilingual. Age of onset determines whether one’s exposure to the two languages is *SIMULTANEOUS*, i.e., two languages from birth (or a very young age), or *SEQUENTIAL*, with exposure to a second language (L2) taking place after significant exposure to the L1 (roughly after 3–4 years of age). Degree of usage facilitates a distinction between *PASSIVE BILINGUALISM*, which describes the ability to comprehend, but not (easily) produce, output in one of the two languages, and *ACTIVE BILINGUALISM*, which entails productive performance abilities and engagement in both languages on a rather wide continuum. Linguistic proficiency also contributes a distinguishing characteristic: a person might be an active bilingual, but with *BALANCED* or *UNBALANCED* performance ability in the two languages. The type of bilingual trajectory invites further distinctions, fueled by the fact that bilingual competence is a dynamic phenomenon that fluctuates throughout the lifespan. The following definition of a heritage bilingual speaker is indicative of how the complex character of language development may lead to differences in the ultimate linguistic attainment of people that may speak the same languages and may share the same age of onset, yet do not share the same trajectory.

A language qualifies as a heritage language if it is a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society. Like the acquisition of a primary language in monolingual situations and the acquisition of two or more languages in situations of societal bilingualism/multilingualism, the heritage language is acquired on the basis of an interaction with naturalistic input and whatever in-born linguistic mechanisms are at play in any instance of child language acquisition. Differently, however, there is the possibility that quantitative and qualitative differences in heritage language input, the introduction and influence of the societal majority language, and differences in literacy and formal education can result in what on the surface seems to be

arrested development of the heritage language or attrition in adult bilingual knowledge. (Rothman, 2009: 156).

Differences between the operationalized definitions for bilingualism are vast. Moreover, being bilingual is not a static characteristic or an ‘on/off’ experience. As we have noted, recent research indicates that when one considers bilingualism as the spectrum of dynamic experiences it is, multiple variables are shown to affect the occurrence and degree of cognitive and neuro-anatomical adaptations (e.g., Bak, 2016b; Bialystok, 2016; Luk & Bialystok, 2013; Li *et al.*, 2014; De Cat, Gusnanto & Serratrice, 2018; Gullifer *et al.*, 2018; Dash *et al.*, 2019; Beatty-Martínez *et al.*, 2019; DeLuca *et al.*, 2019; 2020; Sulpizio, Del Maschio, Fedeli & Abutalebi, 2020b). The elusiveness of bilingual effects, then, could be related, at least partially, to the polysemous nature of the term ‘bilingual’, referring to very different populations across studies. Does a simultaneous bilingual with balanced exposure to two languages have the same (amount of) experience (i.e., in terms of inhibition, control, opportunity for code-switching, actual use, and whatever other factor may be relevant) as a sequential bilingual with limited L2 exposure only in some registers? Can we safely assume that all simultaneous bilinguals are equally comparable in the relevant ways as well? To the extent bilingual experiences matter, if individuals have sufficiently different ones, should we not expect differences in their behavioral outcomes (and neuro-anatomical adaptations) too? If so, might these distinctions contribute to explaining at least some of the non-uniformly attested results across groups from distinct studies, not to mention individuals within the same study?

The heterogeneity of the qualification criteria for bilingualism carries important implications for systematic reviews and meta-analyses (e.g., Adesope, Lavin, Thompson & Ungerleider, 2010; Hilchey & Klein, 2011; de Bruin, Treccani & Della Sala, 2015; Donnelly, Brooks & Homer, 2015; Paap *et al.*, 2015; Lehtonen *et al.*, 2018). Regardless of their conclusion in terms of whether there is enough evidence for consistent bilingual adaptations at the behavioural or brain levels or not, such meta-analyses almost always rely on the original studies’ description of participants’ as being “bilingual”. The caveat is that it is very unlikely that the sets of bilinguals presented in the original studies have the same or even comparable experiences leading to their bilingualism. To give a recent example, Lehtonen *et al.* (2018) are explicit on how they assume the labelling of participants as bilinguals or monolinguals as it appears in the sources, despite the large variation in the definition of bilingualism that these sources assumed (for instance, compare the late bilinguals of Waldie, Badzakova-Trajkov, Milivojevic & Kirk, 2009, who are L1 attriters of Macedonian with L2 English recruited from a monolingual society, to the simultaneous Spanish–Catalan bilinguals of Costa, Hernández and Sebastián-Gallés, 2008, recruited from a bilingual society). Non-uniformity of the bilingual group is not a problem relevant only in the context of meta-analyses, but also in original experimental studies. For example, the bilingual group in D’Souza, Moradzadeh and Wiseheart (2018), who find a musical training advantage but not a bilingual one, involves speakers of English and a second language, the latter being one of 32 languages from different language families. The proficiency of these bilinguals is also quite diverse; nevertheless, fully fluent, active bilinguals and practical bilinguals (i.e., those that reported to be able to carry out conversations fluently, but do not use both languages daily) are placed in the same group. This very same issue, of course, also arises in relation to studies that claim to find bilingual

effects. For instance, in the well-powered study of Brito and Noble (2017), advantageous effects are reported, but the bilinguals (what they call ‘dual-language users’) were classified as such on the basis of a positive answer to a single question, namely “Does the participant speak another language other than English?” (p. 4). Theoretically speaking, a positive answer could entail anything from a fully fluent simultaneous bilingual to a foreign language learner with very limited exposure through instruction.

Thus, in meta-analyses non-uniform groups of people are treated uniformly, being grouped under the rubric ‘bilingual’. These people are indeed described as bilingual in the original studies, but each of these studies usually operates on the basis of ONE established definition per participant group (e.g., simultaneous Spanish–Catalan bilinguals in Catalonia, sequential heritage learners of Russian in the United States, unbalanced Sardinian-Italian bidialectals in Italy, etc). However, when a term is employed in two or more senses WITHIN THE CONTEXT OF ONE SINGLE ARGUMENT, then the argument might ring too close to the fallacy of equivocation. This fallacy occurs when a key notion in an argument is used in an inconsistent or ambiguous way, with one meaning in one part of the argument and another meaning in another part of the argument. The question then becomes more complex, and a binary ‘yes’ or ‘no’ to the question of bilingual effects simply does not suffice. The question becomes: what is it within the profile of groups in terms of linguistic variables that may cause cognitive and neuroanatomical changes to obtain, apparently differentially, and conspire to make individuals and groups distinct?

On the behavioral front, another challenge that has been discussed in relation to meta-analyses comes from the ecological fallacy, which arises when the averages of the participants’ features at the group level (both target and control group) fail to reflect their individual-level characteristics, as argued by Greco, Zangrillo, Biondi-Zoccai and Landoni (2013) on meta-analyses in the field of cardiovascular disease. In light of our discussion of bilingualism as a spectrum of experiential factors, it is important to highlight the obvious: considerable variation is bound to exist at the individual level within and across studies, even in so-called monolingual control groups. It is virtually impossible that different scholars from unique research centers and parts of the world have employed the exact same inclusion criteria for their so-called monolingual and bilingual populations, administered the same background and language proficiency checks to determine ‘monolingual’ and/or ‘bilingual status’, and trimmed the data on the demographic front in an identical or otherwise comparable way. For this reason, it could be the case that meta-analyses and systematic reviews operate on the assumption that they group together similar populations, when in fact they don’t. This heterogeneity may induce some scepticism about the ecological validity of the results.

None of these pitfalls should make us question the value of meta-analyses and systematic reviews as a scientific tool. However, with respect to the topic at hand, the vast heterogeneity that appears to be inherent to populations that are eventually grouped together may explain why different meta-analyses reach contradictory conclusions about the existence of bilingual effects (e.g., Adesope et al., 2010; Lehtonen et al., 2018). It may also explain why some meta-analyses challenge the size and the type of evidence for such effects, while at the same time leaving open the possibility that an effect exists under “very specific AND undetermined circumstances” (Paap et al., 2015; emphasis added). This last view may seem paradoxical, but it is not, if one accepts the aforementioned claim about multi-causality and

forces that work in opposite directions. To repeat, if the various sightings of a bilingual effect are the result of different interactions, there is more than one way of obtaining such an effect. Some ways appear linked to highly specific conditions, because they are found in just a subset of a bigger bilingual population, while at the same time, the contribution of each individual factor (i.e., level of education, proficiency, degree of switching, age of onset of bilingualism, distribution of use of the languages etc.), AND THE POSSIBLE INTERACTIONS among factors remain undetermined. Looking forward then, a collective effort that recognizes that bilingualism is not a categorical variable and seeks to maximize comparability across studies will be in a better position to peel back the layers of the complex questions we seek to answer, a point to which we return below.

Sample size and power

The issue of sample size is perhaps the thorniest one in the context of obtaining reliable evidence for the (non-)existence of bilingual effects. The issue is not restricted to bilingualism research, but pertains to all (or most) psychological research, as using small samples is a general drawback of the field of experimental psychology and cognitive neuroscience (see Brysbaert, 2019, for discussion). Size differences and power variability may explain why some studies find positive evidence, while others do not. More concretely, although numerous studies adduced results that point to the existence of advantageous effects, the effect size of this phenomenon has rightly been questioned. For example, Paap et al. (2015) claim that evidence for bilingual effects often come from small (er) studies, while big studies tend to give null results. While studies published after this observation offer some counterevidence (e.g., Brito & Noble, 2017; Hartanto, Toh & Yang, 2018; De Cat et al., 2018), the original point is a fair one indeed. In this context one wonders what the appropriate sample size should be and what percentage of relevant research meets it.

As Bakker (2015) highlights, if the size threshold for adequate power is $n > 138$ for each group, only 2/86 studies reviewed in Paap et al. (2015) are well-powered; the remaining studies have an average of 35 participants in each group. This is important, because performance in cognitive tasks cannot only be shaped by behavioral experiences such as exposure to more than one language in the course of development. The individual genetic profile also plays a role, as certain genes affect neural activity and consequent performance during cognitive control tasks, while the presence/absence of some behavioral effects may be modulated by prenatal differences in brain morphology (see, for instance, the role of the DRD2 gene, related to dopamine availability in the striatum; Vaughn, Ramos Nuñez, Greene, Munson, Grigorenko & Hernandez, 2016, or the intersubject differences in cognitive control – also across monolinguals and bilinguals – that stem from variability in the anterior cingulate cortex; Del Maschio, Sulpizio, Fedeli, Ramanujan, Ding, Weekes, Cachia & Abutalebi, 2019). Low power increases susceptibility to the ‘individual’ factor, which is a primary suspect for the phantom-like appearance of the bilingual effects. The reason is that in small-scale studies, the impact of individual variation due to (epi)genetic factors, can be particularly impactful, while in well-powered studies, it is increasingly likely to be washed out. This may explain why small studies have been associated with a higher degree of heterogeneity than larger studies (Int’Hout, Ioannidis, Borm & Goeman, 2015).

Sample size is relevant for the credibility and magnitude of the claims one makes. In most fields, the majority of published papers

report statistically significant results, and yet, both the results and the conclusions drawn on their basis are likely to be false (Ioannidis 2005). Size plays a role, because all other factors being equal, a result is more likely to be true in scientific fields that undertake large studies than small ones, as a decrease in size entails a decrease in power (Ioannidis, 2005; Szucs & Ioannidis, 2017). Aiming to put in perspective the $n = 35$ mean size that was mentioned above in relation to the meta-analysis of Paap et al. (2015), we searched PubMed for recent studies that measure behavioral outcomes in the context of the so-called bilingual advantage. The search terms were “bilingual advantage” and “bilingual benefit” and the time window for publication was 01/01/2018–01/08/2018. The only exclusion criterion was the absence of a monolingual control group. Having identified eight relevant studies (table 1), we observe a slight increase in power from the previously reported means: the mean size was $n = 38$ for the bilingual groups and $n = 50$ for the monolingual control groups.

Although sample size matters, it is not a deterministic factor that can guarantee obtaining evidence for or against an effect. To illustrate why this is so, we briefly examine how the factor of sample size interacts with other factors, by discussing some aspects of the two well-powered studies discussed in Paap et al. (2015): Duñabeitia et al. (2014) and Antón et al. (2014). Both studies report results from Spanish-Basque typically developing children. Also, both studies fail to find evidence for bilingual effects (but see later work by Antón, Carreiras & Duñabeitia, 2019 for results that show bilinguals from the very same region outperforming monolinguals on some working memory tasks). Given their (i) power, (ii) meticulous design, and (iii) adequate control measures and careful across-group matching in terms of various indices, it comes as no surprise that Paap et al. (2015) highlight the importance of these two studies and comment that “[they] are noteworthy because the bilinguals acquired both languages early, were highly proficient, and were immersed in a bilingual region” (p. 268).

The linguistic profile presented in Antón et al. (2014) and Duñabeitia et al. (2014) suggests that these children are not simultaneous bilinguals: Spanish was acquired first (0.58 and 0.75 years in Antón et al., 2014 and Duñabeitia et al., 2014 respectively) and Basque well after (2.23 and 2.27 years in Antón et al., 2014 and Duñabeitia et al., 2014 respectively). However, they are clearly active bilinguals insofar they were all attending bilingual schools with a teaching system that grants approximately half of the school time using each of the languages as vehicle for communication. Moreover, they were selected by the authors precisely because of their very high proficiency in both languages. Sample size alone, however, does not guarantee adjudicating between possibilities. And so while these studies are exceptional for their power, the facts related to their highly self-selecting profile for inclusion might only tell us about bilingual effects (or lack thereof) under specific conditions. Our point is that bigger is only better when the sample is populated by the right type of subjects. And what ‘right’ means here can only be solved with an *a priori* complete and unbiased characterization of the multifactorial essence of the bilingual experience.

Defining this right type of subjects is very much an open issue. In certain studies (e.g., Antón et al., 2014; 2019 and Duñabeitia et al., 2014), there is an effort to control for specific critical proxies for bilingual experiences to ensure some consistency, if not relative homogeneity for certain variables such as balanced and high proficiencies in an arguably comparable context, such as

Table 1. Summary of studies on the ‘bilingual advantage’. (B = Bilingual sample size, M = Monolingual sample size, B = M = no difference between monolinguals and bilinguals, B < M = monolingual advantage, B > M = bilingual advantage, OLD = older subsample)

Study	B	n	M	n	Inhibition	Shifting	Updating	Discriminating	Working Memory	Verbal fluency	Conflict resolution	Picture-word identification	Monitoring	Sound contrast sensitivity
Arizmendi et al. (2018)	80	167	B = M	B = M	B = M	B = M	B < M							
Levi (2018)	19	22			B > M									
D'Souza et al. (2018)	72	81	B = M	B = M				B = M						
Pino Escobar et al. (2018)	17	17	B = M						B > M					
Desjardins & Fernández (2018)	19	20	B = M											
Desideri & Bonifacci (2018)	25	25								B > M	B > M	B = M		B > M
Singh et al. (2018)	24	24												
Del Maschio et al. (2018)	44	44			B _(OLD) > M _(OLD)									
Average	38	50												

immersion in fully bilingual societies. At the same time, proficiency or balance may not be the most critical measures to tap into. Proficiency is merely a proxy for how close or distant an internalized grammar X is to the expected, prescriptive norms of X, but no one, at least in linguistics, would claim that a high degree of possible discrepancy between a bilingual's language competence for X and the expected norm of X would entail absence of a comprehensive system for the bilinguals' mental grammar version of language X. If there are two internalized systems in use then, however close or distinct from their corresponding standard norms, we have the makings of competition upon which the mechanisms implicated in conferring bilingual effects should be engaged. Similarly complex is the notion of balance. If the use of the two languages fluctuates throughout the lifespan (e.g., a balanced bilingual education can be succeeded by a working environment that requires the predominant use of one language), an end-state that can be called 'balanced' is probably short-lived and subject to many changes throughout the bilingual speaker's life. More importantly, language (like any other skill) progressively transitions from a heavily controlled process to a far more automated one. It is possible that so-called balanced, simultaneous bilinguals have long-since automated their bilingual language control and receive less practice in top-down cognitive control compared to a sequential bilingual who must suppress a dominant L1 in order to use the L2 (Paap, 2018). Of course, the question remains: if balance and/or proficiency are not the most or only critical measures, what are the factors that can lead to the most robust occurrence of bilingual effects? Decades of research on bilingual cognition have examined a great variety of populations and critical values for key variables have been tested so far, such that there are samples falling into a plethora of categories of bilingual experiences. The outcome, however, has been that proposed theoretical taxonomies do not align with the expected results, and no specific category has been robustly linked to bilingual effects so far. Section 'A roadmap for further work: Designing multi-lab studies' further discusses this with the aim to set a context that could prove fertile for discovering consistent bilingual effects or rule them out completely.

Task effects

It is common to examine cognitive effects of bilingualism through tasks that measure executive functions. Doing so is completely fair, given that the original claims were made on the basis of such task performance differences between monolinguals and bilinguals. However, one cannot ignore that test-retest reliability for such tasks can be (surprisingly) low across the board (see e.g., Karalunas, Bierman & Huang-Pollack, 2016; Chan, Shum, Touloupoulou & Chen, 2008), even in the five most commonly used tasks (see Soveri, Lehtonen, Karlsson, Lukasik, Antfolk & Laine, 2018). The implications of this should not be understated. Indeed, it affects all subfields/studies that rely on such data to support and/or negate specific claims. Thus, we must be cautious in how we interpret evidence related to behavioral effects, or lack thereof, on such tasks. The field of bilingualism would be wise, moving forward, to not rely so heavily on them, if at all, to argue for or against bilingual effects on cognition, given the ubiquitous phantom-like appearance often found in the greater context of executive function task testing.

Low test-retest reliability does not immediately indicate that such tasks are invalid or not entirely fit-for-purpose. There are many extraneous variables that could affect task performance at

any given instance. And so, how do we responsibly explain away the many instances of positive effects? Are they all artefacts? If it turns out to be the case that executive function tasks are simply not reliable enough by their very nature, then the only responsible conclusion would be the neutral one and testing should expand to other domains, going beyond executive functions.

Further complications involve the fact that the construct of executive functions is not as unitary as one may think. Executive functioning involves various components, among them inhibition, switching, attention shifting, and working memory. Even within one of these components, a specific task may target and thus measure different things: for example, testing inhibition might mean testing the ability to inhibit prepotent responses as well as the ability to resist interference by a distractor (Rey-Mermet, Gade & Oberauer, 2018). As a result, an additional contributory factor for the non-replicability of certain findings may be the fact that the instruments used to measure the dependent variable (i.e., executive control) vary from study to study. For one, age of acquisition is known to play a role with respect to which parts of the cognitive system are most affected, with early acquisition favoring switching and late acquisition favoring inhibition (Tao, Marzecová, Taft, Asanowicz & Wodniecka, 2011). If different bilingual trajectories impact the different domains of executive functioning in a variable way, bilingualism research should take into account the interaction between trajectory, the type of task performed, and the subsequent task effects (Cox, Bak, Allerhand, Redmond, Starr, Deary & MacPherson, 2016).

Another important interaction possibly obscuring results is the interaction between task effects and age of testing. Studies that involve both young and older participants have found that older bilinguals are more efficient at inhibiting distracting information than older monolinguals, but the effect may not be seen in the younger sample and/or in all the versions of a task (see Salvatierra & Rosselli, 2010 for the Simon task). Different versions of the same task or different conditions within a task modify the occurrence of an effect. Costa et al. (2008) showed that the bilingual effect can be selectively seen in one version/condition of the task at hand, e.g., affecting the direction of switching (from congruent to incongruent trials or from incongruent to congruent trials) in a conflict resolution task.

Overall, it is important to keep present that both sides of the debate are predicated on the usefulness and appropriateness of the employed tasks. One cannot assume that null or negative results are more reliable than positive ones, or vice versa, if the very nature of the instruments itself contributes to the phantom-like appearance of an effect. We would simply have to concede that more work is needed to understand the variables, including honing in on more reliable methods capable of capturing an overall effect. And in the absence of such methods, the use of several measures or tasks that seemingly tap into the same processes is advised.

Publication bias and the Proteus phenomenon

The current state of the art on the impact of bilingualism on cognition involves several studies that represent seemingly dichotomous sides: one that argues, without denial of the fact that it does not always obtain, in favor of a positive correlation, and one that argues that the obtained evidence has an effect size that is indistinguishable from zero and lacks the consistency of a robust effect. It has not always been this way, however. As de Bruin and Della Sala (2015: 375) put it, "[t]he pattern of supporting versus challenging studies has indeed changed over time. Whereas earlier

studies largely supported a bilingual advantage, recent years (especially 2014) have shown an upsurge in studies challenging this view". It seems that the current balance between studies that report a bilingual effect and those that do not find any is not an accidental one.

Irrespective of the field or the phenomenon at hand, scientific breakthroughs almost always start and progress with positive results; negative results emerge only after a while, possibly as a regression to the mean after an early magnification of the newly found effect (Schooler, 2011). The reason is that there is an initial publication bias that disfavors null or small-size results in the context of a newly explored hypothesis. This naturally occurring cycle often leads to the publication of the most-favorable findings, while at the replication stage, the least-favorable results will likely emerge (Ioannidis & Trikalinos, 2005). This rapid alternation between radically different claims that occurs after a scientific breakthrough has been called the *PROTEUS PHENOMENON* (Ioannidis & Trikalinos, 2005). In this context, the phantom-like appearance of the bilingual effects on cognition – which at the present stage consist of seemingly contradictory results – is the outcome of a time-induced trade-off between an early publication bias that favors positive results and the subsequent Proteus phenomenon.

Sample size and degree of power interact with publication bias in at least two ways. First, small studies are associated with yielding particularly big results (Fanelli, Costas & Ioannidis, 2017). As a matter of fact, small-study effects have been shown to be “the most important source of bias in meta-analysis, which may be the consequence either of selective reporting of results or of genuine differences in study design between small and large studies” (Fanelli et al., 2017: 3717). Second, but related to the previous point, small studies are more likely to be subject to publication bias, especially if they report a small in magnitude negative result: If a researcher completes a very large trial, the result is likely to be published regardless of the outcome, because of the amount of effort involved; however, small negative trials are more likely to remain in the drawer (Lee & Hotopf, 2012).

Relating the two points, it seems that pressure to publish leads to a potential augmentation of the magnitude of the claim in small studies as a compensation for reduced sample size. The complex dynamics behind the publication bias and the Proteus phenomenon may explain why the current literature on the bilingualism effect on cognition involves largely opposite claims, which grant certain positive outcomes to a phantom-like appearance. But one needs to proceed with caution to potentially impulsive shifts in the pendulum inducing a *Zeitgeist* effect in the opposite direction of what is claimed by some to be the same effect originally in the other direction. In other words, we would not want to conclude definitively the opposite of the original claims until there is truly enough solid research to entirely discard the phenomenon.

A roadmap for further work: Designing multi-lab replication studies

The bilingual cognitive effects hypothesis has always been predicated on the proposal that bilingual language control recruits general executive control. However, recent results have questioned the idea of domain-general inhibitory control as a unitary construct. Rey-Mermet et al. (2018) provide compelling evidence that the inhibition measures from 11 established tasks correlate only weakly among each other, calling into question the conceptualization of inhibition as a unitary, psychometric construct. This result casts some doubt on the claim that the experience of bilinguals in

inhibiting one of their languages should consistently lead to enhanced performance in executive function tasks that require inhibition of prepotent responses (e.g., the Stroop task).

In light of the many studies that do find bilingual performance effects, we do not claim that inhibition in the domain of language use does not enhance inhibition in other domains, but that (i) the effect should not be expected to be consistent, and (ii) identifying exactly what mechanisms drive the effect, as others have pointed out, is far from complete. Our aim in this section is thus to provide a multifactorial roadmap for finding the conditions that drive effects and may lead to observing them in the clearest way.

The first factor to take into account is the need for laying out a solid methodology to correctly characterize the intricacies of bilinguals' experience and knowledge. In this line, and considering the bulk of evidence showing reliable effects, one necessarily needs to consider the amount of *OBLIGATORY LANGUAGE SWITCHES* in a bilingual's performance (e.g., through addressing different monolingual interlocutors), the control of which requires frequent engagement of top-down control mechanisms (Blanco-Elorrieta & Pyllkkänen, 2018). To articulate the prediction more clearly, it is possible that the frequent engagement of top-down control processes, which has been explicitly linked to stimulus-driven switching in dense code-switching contexts, may be the key to such effects. Degrees of such top-down processes may condition the likelihood and levels of bilingual effects across individuals and groups (Green & Wei, 2014; Hofweber, Marinis & Treffers-Daller, 2016; Green, 2018). Besides, in addition to the factors already discussed, we would like to argue that studies of bilingual effects should also consider issues related to the languages involved, such as the socio-linguistic dimension, as social prestige may be a proxy for language use in different contexts, as well as the relative typological proximity among the languages, since more closely related varieties that have similar grammars and many cognates could offer fewer opportunities for stimulus-driven code-switching due to high mutual intelligibility. The notion of language proximity is particularly important (Grohmann, 2014, Grohmann & Kambanaros, 2016) and needs experimental evidence to properly adjudicate. After all, it is also possible that closely related varieties require *more* resources for inhibition precisely because it may be harder to suppress a subset of similar representations compared to typologically distant ones (Rothman, 2015).

In the second step of this roadmap, we want to emphasize the importance of collaboration across multiple labs and the use of registered reports, in order to avoid publication biases. If it is the case that the phantom-like appearance of bilingual cognitive effects relates, in part, to idiosyncratic differences in exposure to and use of the languages, then it seems reasonable that these effects would be best tested via multi-lab collaborations. In fact, if multi-lab projects truly take off, the obvious increase in numbers of participants tested under maximally comparable (exactly the same) measures will also address the ubiquitous, yet not easily addressable statistical power issues discussed at length above. While it is true that individual bilinguals even in the same context can vary in how they use their languages in different settings (work, family, etc.), it is of course also the case that trends across groups exist. Geographical happenstance can be a huge plus in terms of helping to control for and thus test variables that may matter for delimiting the types of experiences that give rise to bilingual cognitive effects, while keeping other key factors constant for meaningful comparison across studies. Capitalizing on various geographical sites for data collection via multi-lab projects will also increase diversity of relevant bilingual experiences at the

individual level. Doing so in much larger (combined) samples will provide a greater chance of capturing the precise conditions that lead to bilingual effects, if any, while dealing with potential homogeneity issues that might obtain in large cohort studies when participants are tested under conditions where variability to key, potential factors is reduced (e.g., when tested in a societal bilingual context).

To provide a tangible example, let us imagine a multi-lab collaboration that seeks to understand if indeed some contexts of bilingualism afford a greater opportunity to capture cognitive effects compared to others and capitalizes on one of the languages being kept constant in all locations of testing. Keeping one language constant will form a common basis for linguistic comparison by allowing for the systematic testing of various factors that cluster differentially with it in unique settings. Spanish is a great example, due to its presence across the globe and how it varies in (i) prestige, (ii) languages with which it is in contact, and (iii) tendencies for providing likely opportunities for use. For example, Spanish can be the main societal language or the minority language. In the former situation, it exists under various contexts. For example, in parts of Spain it is definitively the only main societal language, whereas this is not the case in northern regions like Galicia, Basque Country and Catalonia. Even in these bilingual regions, dominance in and patterns of use with Spanish can vary greatly depending on whether a community is more urban or rural. Although Spanish is a prestigious language in all contexts, the other languages are also of high prestige. In Latin America, Spanish exists in a monolingual sense or, like in Spain, it may co-exist in bilingual settings. It is in contact with indigenous languages such as Quechua, Nahuatl and K'iché, and again there is a rural versus urban divide. This divide tends to be more drastic whereby Spanish typically has hegemonic value, even if it is not the main language of a given community, for example in the Andean mountain regions. Spanish is definitively the language of prestige, while indigenous languages vary considerably in terms of acceptance in the mainstream. In Paraguay for example, Guaraní is a co-official language. Even when the other language is held constant as well, say English, the situation can be very different across different communities. Spanish can be a low-prestige language, as in the US, or a high-prestige language, as in the UK. Of course, we cannot completely generalize, since Spanish in the US is not the same depending on region; for example, it is much more prestigious in Miami than in borderland Texas for various historical and political reasons. As mentioned above, language prestige may be a proxy for socio-economic status (SES) and all that this entails. In this panorama of Spanish bilingualism we note that the same language is in contact with many different types of languages, such as agglutinative indigenous languages in Latin America (or Basque in Northern Spain) or other Romance languages (Portuguese, French). Spanish is also one of the most popular second languages studied across the world, from contexts where the main language is a related language, as in Brazil, to contexts across the United States where opportunity for use and out-of-classroom exposure varies significantly.

These factors can help us, by virtue of multi-lab comparisons using the same measures and methods, to fully understand the relative weights of key potential aspects differentiating these groups and individuals in terms of cognitive bilingual effects. There is no shortage of great labs across the globe where Spanish exists as either (one of) the main societal language(s), a minority home language under various SES conditions, or a popular second language. Once a common set of experiments

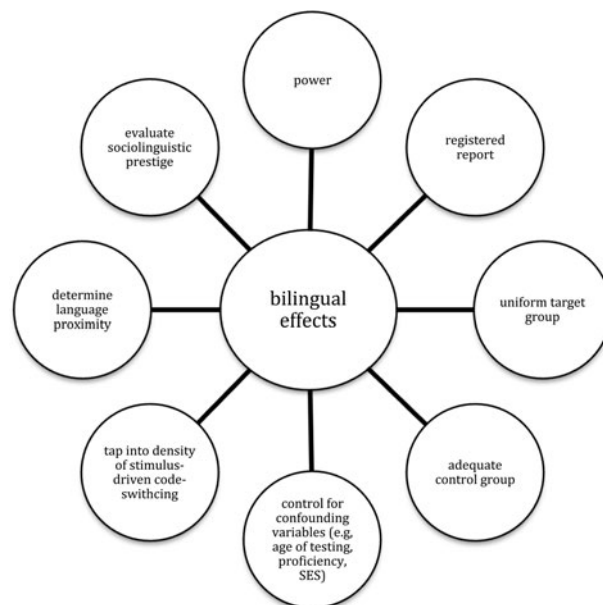


Fig. 2. A summary of critical factors that are relevant for capturing the source and robustness of the bilingual effects

and procedures are agreed upon, and common, exhaustive background measures that can record the information needed to regress over the performance data are identified, all that is needed is the participation of as many labs as possible to capture as much of the spectrum as possible. If we are on the right track, we would expect to see patterns emerge across findings that make the sum worth more than each individual part. With enough labs participating we might be able to uncover with precision which variables in which proportions are more or less likely to result in positive or null effects. Doing so might reveal that there are truly no effects, or alternatively, what the conditions are for effects to obtain. There is a good chance that a large multi-lab endeavor like this one will, no matter what is revealed, be in the best position to make sense of the seemingly contradictory evidence in the literature, by filling gaps between studies that are, to date, not accounted for or properly considered.

Although no study can eliminate all the confounding variables that may drive the conditions that determine bilingual effects (Bak, 2016b), including the ‘individual factor’ mentioned above, we may summarize the methodological issues in the following way: A study will have a greater likelihood to uncover the origin of such effects if (i) it is a well-powered one that (ii) involves multi-lab collaboration, (iii) uses bilinguals of the same type with a nuanced perspective of bilingualism in mind, (iv) employs ADEQUATE comparison groups for baseline, (v) proceeds on the basis of registered reports, (vi) controls for various critical confounding variables, such as age of onset, age of testing, SES, and language proficiency, (vii) tests the impact of frequency of stimulus-driven code-switching, (viii) considers the social dimension of language use, (ix) takes language proximity into account, and (x) makes use of different tasks to approach one construct. We specifically hypothesize that the effects would be seen at their clearest when simultaneous or early active bilinguals that speak typologically distant languages are tested, in a dense, stimulus-driven code-switching context and in a sociolinguistically balanced setting in terms of the prestige ascribed to the two languages. **Figure 2** summarizes the relevant critical factors/measures.

Outlook

Herein, we have discussed the phantom-like appearance of bilingual effects on cognition by approaching them as the multi-causal outcome of several factors. Such effects, advantageous and not, are gradable, dynamic phenomena, whose different manifestations may have a different origin from case to case, depending on the individual characteristics at play. We have laid out a roadmap for future work that sidesteps contentious debate and lays out a set of common procedures, the following of which will increase our collective chances at revealing the origin of robust bilingual effects, if existent. We discussed several methodological points that should be of interest to researchers aiming to understand bilingual effects, regardless of where they think the cards will ultimately fall in the debate that currently surrounds this topic. Based on a careful evaluation of arguments across the aisles as well as a review of various critical measures, our overall prediction is that that bilingual effects would be seen at their clearest when testing actively engaged bilinguals on a continuum, potentially the most under idealized situations of engaging the mechanisms involved to the max, for example, in those that speak typologically distant languages, in a dense stimulus-driven code-switching context, and in a sociolinguistically balanced setting in terms of the prestige ascribed to the two languages.

Bilingualism represents a distinctive way to investigate how brain and behavior affect one another, and the role environmental factors play in modulating this relationship. We have suggested that research should continue in a modified way, because we are ultimately interested in capturing the dynamic interplay between the various factors identified above: a research objective that is currently at the core of cognitive neuroscience. The presence of largely contradictory findings across small- and large-scale studies in the current literature suggests that the field has reached a level of maturity beyond the initial alternation of positive or negative results. This may pave the way for a much-needed change of focus: from debating the absence/presence of a uniform bilingual effect on the anticipation of big differences and deterministic factors to examining the interactions of variables that may drive even marginal differences and how these may vary across studies, tasks, populations, and types of bilingual trajectories.

References

- Adesope AA, Lavin T, Thompson T and Ungerleider C (2010) A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research* **80**, 207–245.
- Aisen PS, Cummings J, Jack CR, Jr, Morris JC, Sperling R, Frölich L, Jones RW, Dowsett SA, Matthews BR, Raskin J, Scheltens P and Dubois B (2017) On the path to 2025: understanding the Alzheimer's disease continuum. *Alzheimer's Research & Therapy* **9**, 60. <https://doi.org/10.1186/s13195-017-0283-5>
- Alladi S, Bak TH, Duggirala V, Surampudi B, Shailaja M, Shukla AK, Chaudhuri JR and Kaul S (2013) Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology* **81**, 1938–1944. doi:10.1212/01.wnl.0000436620.33155.a4
- Antón E, Duñabeitia JA, Estévez A, Hernández JA, Castillo A, Fuentes LJ, Davidson DJ and Carreiras M (2014) Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Language Sciences* **5**, 398.
- Antón E, Carreiras M and Duñabeitia JA (2019) The impact of bilingualism on executive functions and working memory in young adults. *PLoS ONE* **14**, e0206770. <https://doi.org/10.1371/journal.pone.0206770>
- Arizmendi GD, Alt M, Gray S, Hogan TP, Green S and Cowan N (2018) Do bilingual children have an executive function advantage? Results from inhibition, shifting, and updating tasks. *Language, Speech, and Hearing Services in Schools* **49**, 356–378.
- Bak TH (2016a) Cooking pasta in La Paz. *Linguistic Approaches to Bilingualism* **5**, 1–19.
- Bak TH (2016b) The impact of bilingualism on cognitive ageing and dementia: finding a path through a forest of confounding variables. *Linguistic Approaches to Bilingualism* **6**, 205–226.
- Bakker M (2015) Power problems: $n > 138$. *Cortex* **3**, 367–368.
- Baum S and Titone D (2014) Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics* **35**, 857–894.
- Beatty-Martínez AL, Navarro-Torres CA, Dussias PE, Bajo MT, Guzzardo Tamargo RE and Kroll JF (2019) Interactional Context Mediates the Consequences of Bilingualism for Language and Cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xlm0000770>
- Belluz J and Hoffman S (2015) The one chart you need to understand any health study. <https://www.vox.com/2015/1/5/7482871/types-of-study-design>
- Bialystok E (2011) Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology* **65**, 229–235.
- Bialystok E (2016) The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism* **6**, 517–534. <http://doi.org/10.1075/lab.15040.bia>
- Bialystok E (2017) The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin* **143**, 233–262. <http://doi.org/10.1037/bul0000099>
- Bialystok E, Craik FIM and Freedman M (2007) Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia* **45**, 459–464.
- Bialystok E, Craik FIM and Luk G (2008) Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **34**, 859–987.
- Bialystok E and DePape AM (2009) Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance* **35**, 565–574.
- Blom E, Boerma T, Bosma E, Cornips L and Everaert E (2017) Cognitive Advantages of Bilingual Children in Different Sociolinguistic Contexts. *Frontiers in Psychology* **8**: 552.
- Brito NH and Noble KG (2017) The independent and interacting effects of socioeconomic status and dual-language use on brain structure and cognition. *Developmental Science* e12688.
- Blanco-Elorrieta E and Pykkänen L (2018) Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences* **22**, 1117–1122.
- Bloomfield L (1933) *Language*. Holt: New York.
- Brybaert M (2019) How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition* **2**, 16.
- Burgaleta M, Sanjuán A, Ventura-Campos N, Sebastián-Galles N and Ávila C (2016) Bilingualism at the core of the brain: Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *NeuroImage* **125**, 437–445.
- Chan RCK, Shum D, Touloupoulou T and Chen EYH (2008) Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology* **23**, 201–216.
- Chertkow H, Whitehead V, Phillips N, Wolfson C, Atherton J and Bergman H (2010) Multilingualism (but not always bilingualism) delays the onset of Alzheimer Disease: Evidence from a bilingual community. *Alzheimer Disease & Associated Disorders* **24**, 118–125. doi:10.1097/wad.0b013e3181ca1221
- Costa A and Sebastián-Gallés N (2014) How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience* **15**, 336–345.
- Costa A, Hernández M and Sebastián-Gallés N (2008) Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition* **106**, 59–86.
- Costa A, Hernández M, Costa-Faidella J and Sebastián-Gallés N (2009) On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition* **113**, 135–149.
- Costumero V, Rodríguez-Pujadas A, Fuentes-Claramonte P and Ávila C (2015) How bilingualism shapes the functional architecture of the brain: A study on executive control in early bilinguals and monolinguals. *Human Brain Mapping* **36**, 5101–5112. <http://doi.org/10.1002/hbm.22996>

- Cox SR, Bak TH, Allerhand M, Redmond P, Starr JM, Deary IJ and MacPherson SE (2016) Bilingualism, social cognition and executive functions: A tale of chickens and eggs. *Neuropsychologia* **91**, 299–306.
- Craik FIM, Bialystok E and Freedman M (2010) Delaying the onset of Alzheimer disease: Bilingualism as a form of cognitive reserve. *Neurology* **75**, 1726–1729.
- Dash T, Berroir P, Joannette P and Ansaldo AI (2019) Alerting, orienting, and executive control: the effect of bilingualism and age on the subcomponents of attention. *Frontiers in Neurology* **10**: 1122. doi: 10.3389/fneur.2019.01122
- Debarton U, Sperduti M, Di Rienzo F and Guillot A (2014) Experts bodies, experts minds: How physical and mental training shape the brain. *Frontiers in Human Neuroscience* **8**. doi:10.3389/fnhum.2014.00280
- De Baene W, Duyck W, Brass M and Carreiras M (2015) Brain Circuit for Cognitive Control Is Shared by Task and Language Switching. *Journal of Cognitive Neuroscience* **27**, 1752–1765. <http://doi.org/10.1162/jocn>
- de Bruin A and Della Sala S (2015) The decline effect: How initially strong results tend to decrease over time. *Cortex* **73**, 375–377.
- de Bruin A, Treccani B and Della Sala S (2015) Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science* **99**–107.
- De Cat C, Gusnanto A and Serratrice L (2018) Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition* **40**, 119–151.
- Del Maschio N, Sulpizio S, Gallo F, Fedeli D, Weekes BS and Abutalebi J (2018) Neuroplasticity across the lifespan and aging effects in bilinguals and monolinguals. *Brain and Cognition* **125**, 118–126.
- Del Maschio N, Sulpizio S, Fedeli D, Ramanujan K, Ding G, Weekes BS, Cachia A and Abutalebi J (2019) ACC sulcal patterns and their modulation on cognitive control efficiency across lifespan: A neuroanatomical study on bilinguals and monolinguals. *Cerebral Cortex* **29**, 3091–3101. doi: 10.1093/cercor/bhy175.
- DeLuca V, Rothman J and Pliatsikas C (2018) Linguistic immersion and structural effects on the bilingual brain: a longitudinal study. *Bilingualism: Language and Cognition* **22**, 1160–1175. doi:10.1017/S1366728918000883
- DeLuca V, Rothman J, Bialystok E and Pliatsikas C (2019) Redefining bilingualism: A spectrum of experience that differentially affect brain structure and function. *Proceedings of the National Academy of Science (PNAS)* **116**, 7565–7574.
- DeLuca V, Rothman J, Bialystok E and Pliatsikas C (2020) Duration and extent of bilingual experience modulate neurocognitive outcomes. *NeuroImage* **204**: 116222. doi: 10.1016/j.neuroimage.2019.116222
- Desideri L and Bonifacci P (2018) Verbal and nonverbal anticipatory mechanisms in bilinguals. *Journal of Psycholinguistic Research* **47**, 719–739.
- Desjardins JL and Fernandez F (2018) Performance on auditory and visual tasks of inhibition in English monolingual and Spanish–English bilingual adults: Do bilinguals have a cognitive advantage? *Journal of Speech, Language, and Hearing Research* **61**, 410–419.
- Donnelly S, Brooks PJ and Homer BD (2015) Examining the bilingual advantage on conflict resolution tasks: A meta-analysis. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- D'Souza AA, Moradzadeh L and Wiseheart M (2018) Musical training, bilingualism, and executive function: Working memory and inhibitory control. *Cognitive Research: Principles and Implications* **3**, 11.
- Duñabeitia JA, Hernández JA, Antón E, Macizo P, Estévez A, Fuentes LJ and Carreiras M (2014) The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology* **61**, 234–251.
- Duñabeitia JA and Carreiras M (2015) The bilingual advantage: Acta est fabula? *Cortex* **73**, 371–372.
- Edwards J (2004) Foundations of bilingualism. In T Bhatia and W Ritchie (eds), *The Handbook of Bilingualism*. Oxford: Blackwell, pp. 7–31.
- Fanelli D, Costas R and Ioannidis JPA (2017) Meta-assessment of bias in science. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 3714–3719.
- Filippi R, Morris J, Richardson FM, Bright P, Thomas MSC, Karmiloff-Smith A and Marian V (2015) Bilingual children show an advantage in controlling verbal interference during spoken language comprehension. *Bilingualism (Cambridge English)* **18**, 490–501.
- Fuchs E and Flugge G (2014) Adult Neuroplasticity: More than 40 years of research. *Neural Plasticity* **541870**. doi: <http://dx.doi.org/10.1155/2014/541870>
- García-Pentón L, Fernández García Y, Costello B, Duñabeitia JA and Carreiras M (2016) “Hazy” or “jumbled”? Putting together the pieces of the bilingual puzzle. *Language, Cognition and Neuroscience* **31**, 353–360.
- Greco T, Zangrillo A, Biondi-Zoccai G and Landoni G (2013) Meta-analysis: pitfalls and hints. *Heart Lung and Vessels* **5**, 219–225.
- Green D (2018) Language control and code switching. *Languages* **3**, 8. doi: <https://doi.org/10.3390/languages3020008>
- Green DW and Abutalebi J (2013) Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology* **25**, 515–530.
- Green DW and Wei L (2014) A control process model of code-switching. *Language, Cognition and Neuroscience* **29**, 499–511.
- Grohmann KK (2014) Towards comparative bilingualism. *Linguistic Approaches Bilingualism* **4**, 336–341.
- Grohmann KK and Kambanaros M (2016) The Gradient of Multilingualism in Typical and Impaired Language Development: Positioning Bilingualism within Comparative Bilingualism. *Frontiers in Psychology* **7**: 37. doi: 10.3389/fpsyg.2016.00037
- Guzmán-Vélez E and Tranel D (2015) Does bilingualism contribute to cognitive reserve? Cognitive and neural perspectives. *Neuropsychology* **29**, 139–150.
- Gullifer JW, Chai XJ, Whitford V, Pivneva I, Baum S, Klein D and Titone D (2018) Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia* **117**, 123–134. doi: <http://doi.org/10.1016/j.neuropsychologia.2018.04.037>
- Hartanto A, Toh WX and Yang H (2018) Bilingualism narrows socioeconomic disparities in executive functions and self-regulatory behaviors during early childhood: Evidence from the early childhood longitudinal study. *Child Development* **90**, 1215–1235. doi.org/10.1111/cdev.13032
- Hilchey MD and Klein RM (2011) Are there bilingual advantages on non-linguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review* **18**, 625–658.
- Hofweber J, Marinis T and Treffers-Daller J (2016) Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism* **6**, 648–668. doi: doi.org/10.1075/lab.15052.hof
- Int'Hout J, Ioannidis JPA, Borm GF and Goeman JJ (2015) Small studies are more heterogeneous than large ones: a meta-meta-analysis. *Journal of Clinical Epidemiology* **68**, 860–869.
- Ioannidis JPA (2005) Why most published research findings are false. *PLoS Medicine* **2**, e12.
- Ioannidis JPA (2016) The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly* **94**, 485–514. doi:10.1111/1468-0009.12210
- Ioannidis JPA and Trikalinos TA (2005) Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology* **58**, 543–549.
- Jeon H-A, Kuhl U and Friederici AD (2019) Mathematical expertise modulates the architecture of dorsal and cortico-thalamic white matter tracts. *Scientific Reports* **9**, 6825. doi:10.1038/s41598-019-43400-6
- Karalunas SL, Bierman KL and Huang-Pollack CL (2016) Test-retest reliability and measurement invariance of executive function tasks in young children with and without ADHD. *Journal of Attention Disorders* **1087054715627488**. doi: doi.org/10.1177/1087054715627488.
- Kroll J and Bialystok E (2013) Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology* **25**, 497–514. DOI: 10.1080/20445911.2013.799170
- Kroll J and Chiarello C (2016) Language experience and the brain: variability, neuroplasticity, and bilingualism. *Language, Cognition and Neuroscience*, **31**, 345–348, DOI: 10.1080/23273798.2015.1086009
- Lauchlan F, Parisi M and Fadda R (2013) Bilingualism in Sardinia and Scotland: Exploring the cognitive benefits of speaking a ‘minority’ language. *International Journal of Bilingualism* **17**, 43–56.
- Lawton DM, Gasquoine PG and Weimer AA (2015) Age of dementia diagnosis in community dwelling bilingual and monolingual Hispanic Americans. *Cortex* **66**, 141–145.

- Lee W and Hotopf M** (2012) Critical appraisal. Reviewing scientific evidence and reading academic papers. In P Wright, J Stern and M Phelan (eds), *Core Psychiatry*. London: Saunders, pp. 130–142.
- Lehtonen M, Soveri A, Laine A, Järvenpää J, de Bruin A and Antfolk J** (2018) Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin* **144**, 394–425.
- Levi SV** (2018) Another bilingual advantage? Perception of talker-voice information. *Biling (Camb Engl)* **21**, 523–536.
- Li P, Legault J and Litcofsky KA** (2014) Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex* **58**, 301–324.
- Lieberson S** (1991) Small n's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases. *Social Forces* **70**, 307–320.
- Linnavalli T, Putkinen V, Lipsanen J, Huotilainen M and Tervaniemi M** (2018) Music playschool enhances children's linguistic skills. *Scientific Reports* **8**, 8767.
- Luk G and Bialystok E** (2013) Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology* **25**, 605–621. <http://doi.org/10.1080/20445911.2013.795574>
- Luk G, Anderson JA, Craik FI, Grady C and Bialystok E** (2010) Distinct neural correlates for two types of inhibition in bilinguals: response inhibition versus interference suppression. *Brain and Cognition* **74**, 347–57.
- Luk G, Bialystok E, Craik FIM, Grady CL** (2011) Lifelong bilingualism maintains white matter integrity in older adults. *Journal of Neuroscience* **31**, 16808–16813.
- Maguire EA, Burgess N, Donnett JG, Frackowiak RSJ, Frith CD, O'Keefe J** (1998) Knowing where and getting there: a human navigation network. *Science* **280**: 921–924.
- Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RS, Frith CD** (2000) Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 4398–4403. doi:10.1073/pnas.070039597
- Masouleh SK, Eickhoff SB, Hoffstaedter F, Genon S and Alzheimer's Disease Neuroimaging Initiative.** (2019) Empirical examination of the replicability of associations between brain structure and psychological variables. *Elife* **8**, e43464.
- Morton JB and Harper SN** (2007) What did Simon say? Revisiting the bilingual advantage. *Developmental Science* **10**, 719–726.
- Nichols ES, Wild CJ, Stojanoski B, Battista ME and Owen AM** (2020) Bilingualism affords no general cognitive advantages: A population study of executive function in 11,000 People. *Psychological Science*, 1–20. <https://doi.org/10.1177/0956797620903113>
- Paap KR** (2018) Bilingualism in cognitive science: the characteristics and consequences of bilingual language control. In A De Houwer and L Ortega (eds), *The Cambridge Handbook of Bilingualism*. Cambridge: Cambridge University Press, pp. 435–465.
- Paap KR and Greenberg ZI** (2013) There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology* **66**, 232–258.
- Paap KR, Johnson HA and Sawi O** (2015) Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* **69**, 265–278.
- Papathodorou S** (2019) Umbrella reviews: what they are and why we need them. *European Journal of Epidemiology* **34**, 543–546.
- Parker Jones O, Green DW, Grogan A, Pliatsikas C, Filippopolitis K, Ali N, Lee H.L, Ramsden S, Gazarian K, Prejawa S, Seghier ML and Price CJ** (2012) Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex* **22**, 892–902. <http://doi.org/10.1093/cercor/bhr161>
- Perani D and Abutalebi J** (2015) Bilingualism, dementia, cognitive and neural reserve. *Current Opinion in Neurology* **28**, 618–625.
- Pino Escobar G, Kalashnikova M and Escudero P** (2018) Vocabulary matters! The relationship between verbal fluency and measures of inhibitory control in monolingual and bilingual children. *Journal of Experimental Child Psychology* **170**, 177–189.
- Pliatsikas C** (2019a) Understanding structural plasticity in the bilingual brain: The Dynamic Restructuring Model. *Bilingualism: Language and Cognition* 1–13.
- Pliatsikas C** (2019b) Multilingualism and brain plasticity. In JW Schweiter (ed), *The Handbook of the Neuroscience of Multilingualism*. Wiley Blackwell, pp. 230–251.
- Pliatsikas C and Luk G** (2016) Executive control in bilinguals: a concise review on fMRI studies. *Bilingualism: Language and Cognition* **19**, 699–705.
- Preische O, Schultz SA, Apel A et al.** (2019) Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease. *Nature Medicine* **25**, 277–283. <https://doi.org/10.1038/s41591-018-0304-3>
- Rey-Mermet A, Gade M and Oberauer K** (2018) Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **44**, 501–526.
- Ross J and Melinger A** (2016) Bilingual advantage, bidialectal advantage or neither? Comparing performance across three tests of executive function in middle childhood. *Developmental Science* **20**, e12405.
- Rothman J** (2009) Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* **13**, 155–163.
- Rothman J** (2015) Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition*. doi:10.1017/S136672891300059X.
- Saari P, Burunat I, Brattico E and Toiviainen P** (2018) Decoding musical training from dynamic processing of musical features in the brain. *Scientific Reports* **8**, 708. doi:10.1038/s41598-018-19177-5
- Salvatierra JL and Rosselli M** (2010) The effect of bilingualism and age on inhibitory control. *The International Journal of Bilingualism* **15**, 26–37.
- Schooler J** (2011) Unpublished results hide the decline effect. *Nature* **470**, 437.
- Singh L, Fu CSL, Wen Tay Z and Michnick Golinkoff R** (2018) Novel word learning in bilingual and monolingual infants: Evidence for a bilingual advantage. *Child Development* **89**, e183–e198.
- Soveri A, Lehtonen M, Karlsson LC, Lukasik K, Antfolk J and Laine M** (2018) Test-retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology: Adult* **25**, 155–165.
- Sulpizio S, Del Maschio N, Del Mauro G, Fedeli D and Abutalebi J** (2020a) Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *NeuroImage* **205**, 116306.
- Sulpizio S, Del Maschio N, Fedeli D and Abutalebi J** (2020b) Bilingual language processing: A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews* **108**, 834–853.
- Szucs D and Ioannidis JPA** (2017) Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology* **15**, e2000797.
- Tao L, Marzecová A, Taft M, Asanowicz D and Wodniecka Z** (2011) The efficiency of attentional networks in early and late bilinguals: the role of age of acquisition. *Frontiers in Psychology* **2**, 123.
- Valian V** (2015) Bilingualism and cognition. *Bilingualism: Language and Cognition* **18**, 3–24.
- van den Noort M, Vermeire K, Bosch P, Staudte H, Krajenbrink T, Jaswetz L, Struys E, Yeo S, Barisch P, Perriard B, Lee SH and Lim S** (2019) A systematic review on the possible relationship between bilingualism, cognitive decline, and the onset of dementia. *Behavioral Sciences* **9**, 81. doi:10.3390/bs9070081
- Vaughn KA, Ramos Nuñez AI, Greene MR, Munson BA, Grigorenko EL and Hernandez AE** (2016) Individual differences in the bilingual brain: The role of language background and DRD2 genotype in verbal and non-verbal cognitive control. *Journal of Neurolinguistics* **40**, 112–127.
- Waldie K, Badzakova-Trajkov G, Milivojevic B and Kirk I** (2009) Neural activity during Stroop colour-word task performance in late proficient bilinguals: A functional magnetic resonance imaging study. *Psychology & Neuroscience* **2**, 125–136.
- Yeung CM, St. John PD, Menec V and Tyas SL** (2014) Is bilingualism associated with a lower risk of dementia in community-living older adults? *Cross-sectional and Prospective Analyses. Alzheimer Disease & Associated Disorders* **28**, 326–332.