
Reviewed by Kenneth R. Paap, San Francisco State University

The focus of Bilingualism, executive function, and beyond is the hypothesis that managing two languages requires domain-general executive functioning (EF) and that this ubiquitous practice leads to an enhancement of cognitive control. This hypothesis was first challenged in the context of a failed replication by Morton and Harper in 2007, and the ensuing debate continued to pick up steam for fifteen years and counting. The dominant view expressed throughout the book is that bilinguals do adapt to the demands of managing multiple languages and that these adaptations affect and enhance general control mechanisms. However, each bilingual’s experience is distinctive, as determined by the specific languages acquired, the age of acquisition, the proficiency attained, the amount of use, and the pattern of use. The latter emphasizes a distinction between single-language contexts, where a single language dominates in each context (e.g. English at school, Spanish at home), and dual-language contexts, where there is frequent language switching depending upon changes in conversational partners or topics. With the further refinement and adoption of the theoretical framework that general cognitive control is best viewed as a set of interrelated component processes, one comes to appreciate that different types of bilinguals will recruit and strengthen different component processes of EF. Thus, another major theme of the book is that the hypothesis of a bilingual advantage in EF (the BILINGUAL ADVANTAGE HYPOTHESIS) cannot be tested with a simple comparison between bilinguals and monolinguals but must take the type of bilingual into account; this requires advances in the methods used to measure different facets of bilingualism and in the tasks used to measure different components of EF. The book does a good job of laying out this agenda, although there is no attempt to integrate the separate
chapters into something like a continuous narrative that tells this story. This is not surprising given the book’s origin, as briefly described next.

The book’s genesis was a 2015 NSF-sponsored workshop held at the CUNY Graduate Center. Eleven papers from that workshop were published in 2016 as a special issue of *Linguistic Approaches to Bilingualism*. They constitute a little more than half of the contents of this volume. Nine posters from the workshop were written up as articles and form the rest of the book. Unfortunately, the time between the development of these papers and their appearance in this book means that the book was not able to benefit from advances in research and theory over the intervening years. During this time the vigorous debate about the bilingual advantage hypothesis has grown more intense, as new empirical studies and reviews seem to appear by the dozen every month.

A potential weakness of the book is that it will mislead a new generation of researchers into adopting the view that evidence has confirmed the hypothesis that bilingualism enhances components of general EF. In the remainder of this review I make the case that the narrative is too rosy and incomplete. The critique is organized with respect to four issues: (i) new meta-analyses paint a very pessimistic picture for the bilingual advantage hypothesis, (ii) the presumption that domain-general executive control exists may be false, (iii) despite intensive effort, the necessary and sufficient conditions for generating bilingual advantages are nowhere in sight, and (iv) Ellen Bialystok’s (e.g. 2016) program for separating signal from noise took some dubious steps and has not been successful.

**Meta-analyses do not support the hypothesis.** The book presents a distorted and incomplete view of the evidence because it fails to take into account the recent march of formal meta-analyses (von Bastian et al. 2017, Lehtonen et al. 2018, Donnelly et al. 2019, Paap 2019, Gunnerud et al. 2020, Lowe et al. 2021, Monnier et al. 2021) that cast the relevant research in a new light. A broad summary is that overall effect sizes for all components of EF are very small and not distinguishable from zero when corrected for publication bias. Many of the meta-analyses include tests for many different moderators, including different types of bilingual experience and different measures of EF. The only mention of any of these is in the introductory chapter by the editors (Irina Sekerina, Lauren Spradlin, and Virginia Valian), who characterize the results of Lehtonen et al. 2018 (and an earlier version of Donnelly et al. 2019) in much the same way as I just did. This is critical, because if zero is a good estimate of the overall effect size, then any special subset of bilinguals who consistently show bilingual advantages must be countered by a subset of bilinguals who show a disadvantage that cancels out the advantage. The remaining mathematical possibility consistent with overall effect sizes of about zero is that there is no signal at all—that no set of bilingual experiences systematically enhances domain-general control. However, the editors are not perturbed by the extremely small and rare positive effects. Rather, they assert that ‘[b]y carefully comparing groups that do and do not have bilingual advantages, it should be possible to determine which bilingual backgrounds are either necessary, sufficient, or necessary and sufficient. That, in turn, should make it possible to investigate the underlying mechanisms behind advantages on executive function tasks’ (3). Despite the fact that most of the book’s authors endorse this research strategy, nowhere in the rest of the book is there a specific hypothesis that A, B, and C may be the sufficient conditions for producing bilingual advantages that will consistently replicate across labs. This is not to say that there are not rampant descriptions of the conditions producing bilingual advantages in many specific studies, together with rationalizations for each obtained pattern of results.

The views regarding when and where inconsistencies occur vary across the chapter authors. Veronica Whitford and Gigi Luk’s Ch. 5 observes that ‘[t]he extant literature on this topic has reported a mixed pattern of positive and null results, depending on the age groups tested, experimental tasks, and locations of research’ (68). Likewise, Klara Marton’s abstract (Ch. 17) opens with the observation that ‘[w]hile findings on the bilingual advantage in adults are mixed the data from children are more consistent’ (265). In contrast, Virginia Valian’s Ch. 15 begins with the observation that inconsistency in tests for bilingual advantages ‘is the rule not the exception’ across children, young adults, and older adults. Valian’s view is a refreshing and correct departure from repeating the unsubstantiated lore that advantages are prevalent and consistent in
children and older adults. In fact, the evidence observed with children, from systematic reviews and meta-analyses, shows small and inconsistent advantages at best. Paap 2019 reported that only three of the thirty comparisons using children in the range of six to fifteen years old produced significant bilingual advantages in nonverbal interference scores (assumed by many to measure inhibitory control). Donnelly et al.’s (2019) meta-analysis of global reaction time (RT) and interference scores with participants 4.5 years and older reported a very small bilingual advantage that was not significantly different from zero when corrected for publication bias. Furthermore, no bilingual advantages in nonverbal interference tasks have been shown in large-scale studies with highly proficient bilingual children who live in language communities where language switching occurs regularly (Antón et al. 2014, Duñabeitia et al. 2014). Although Bialystok (2017:238) dismisses these results because they ‘examine an unusually large age range without convincing control over the role of age in performance’, the results in these studies were analyzed in separate and narrow age bands and showed no evidence that age or years of bilingual experience matter.

Adding more weight to the conclusion that bilingual advantages do not consistently or significantly occur in children is the recent meta-analysis of 583 effect sizes by Gunnerud et al. (2020), which showed a tiny overall effect size of $g = 0.06$. Gennerud et al.’s meta-analysis was then echoed by a meta-analysis by Lowe et al. (2021) that covered ten components of EF and also restricted its focus to children. For the general construct of EF, the overall effect size on EF was very small, $g = 0.08$. When the results were corrected for publication bias, the overall effect was now slightly negative, $g = -0.04$. A distinctive aspect of the Lowe et al. meta-analysis was the inclusion of study quality (e.g. matching language groups on potential confounding variables) as a mediator. They report that as study quality increased, effect-size magnitude decreased! It seems that better controls make the advantage more elusive rather than more easily captured. Lowe et al. (2021) concluded that bilingual advantages in children are related not to language status, but to a variety of unmeasured and uncontrolled factors. Finally, there is the curious outcome that all three meta-analyses that included ‘lab’ as a potential mediator found that only one lab consistently produces bilingual advantages in EF.

THOMAS BAK, in Ch. 6, espouses a belief about the empirical landscape (or seascape) that is at odds with the meta-analyses. He states that the difference between monolingual and bilingual groups ‘has a remarkably similar pattern across various studies. It is like a mountain chain submerged under water, hidden and exposed in turn by rising and falling tides nevertheless maintaining its shape. The highest peak of this chain are executive functions … those connected with inhibition, monitoring, and switching’ (83). If this were true, then meta-analyses would show consistent effect sizes for measures of EF. They do not.

ANNE L. BEATTY-MARTÍNEZ and PAOLA E. DUSSIAS (Ch. 4) advocate for bilingual advantages in monitoring as well as inhibition. Hilchey and Klein (2011) were the first to highlight the consistent bilingual advantage in global RT in the spatial Stroop and flanker tasks and hypothesized that it may reflect a better ability to manage attention to a complex set of rapidly changing task demands (see also Paap, Anders-Jefferson, et al. 2020). However, Paap and Greenberg (2013) reported no advantages on ten different measures of monitoring. This foreshadowed Hilchey, Saint-Aubin, and Klein’s 2015 updated review: they reported that for young adults the consistent advantages observed in their 2011 review were canceled out by consistent disadvantages post-2011. Furthermore, both Paap’s 2019 meta-analysis and Lehtonen et al.’s (2018) meta-analyses of the monitoring ‘advantage’ as reflected in global RT yielded mean effect sizes that were not distinguishable from zero. Thus Ch. 4, like many others in this volume, fails to consider the flurry of recent meta-analyses that consistently question the existence of bilingual advantages.

THE PRESUMPTION OF DOMAIN-GENERAL EXECUTIVE CONTROL MAY BE FALSE. The bilingual advantage hypothesis is predicated on the assumption that certain types of bilingual language control recruit and enhance domain-general components of EF. If this assumption is false, then the basis for predicting bilingual advantages vanishes. The chapters by Beatty-Martínez and Dussias (Ch. 4) and Whitford and Luk (Ch. 5) both open with good introductions to the overwhelm-
ing and uncontested hypothesis that even when bilinguals intend to speak a single target language, the non-target language is partially activated. Beatty-Martínez and Dussias observe that this parallel activation rarely leads to unintended intrusions, that bilinguals need to regulate the competition, and that the selection mechanism is inhibition. Likewise, Whitford and Luk assert that bilinguals ‘must habitually recruit their domain general executive-function capacities to resolve the conflict’ (70). But as Paap et al. (2019:76) have pointed out, there are several possible adaptations to this competition: (1) the non-target lexical competitors are inhibited via a general inhibitory control mechanism, (2) the non-target lexical competitors are inhibited via a specialized mechanism within a word identification system, (3) the target candidates are up regulated, or (4) no conflict resolution mechanism is employed at the level of lexical representations. If the last is true, then bilinguals (5) may live with the competition and the occasional unintended intrusion from the other language, (6) employ domain-general response suppression, or (7) rely on a specialized mechanism for articulatory suppression. Some of these options could occur in combination, but logically only when (1) or (6) are involved could bilingual language-control transfer to and lead to bilingual advantages in nonverbal interference tasks requiring manual responses.

The underlying theory for bilingual advantages in EF is that domain-general control is recruited to resolve competition between the two languages. As illustrated in the previous paragraph, this does not follow inevitably from the presence of competition. Consider next the possibility that control is recruited, but that it is specific to the language-processing system and is not domain-general (Paap et al. 2019). Anna Wolleb, Antonella Sorace, and Marit Vestergaard (Ch. 10) chime in on the issue and ask whether language control and domain-general EF are independent or shared processes. If they are separate processes, then there would be no reason to expect far-transfer from bilingualism to nonverbal measures of EF. Such questions do not have yes/no answers, and the study they focus on (involving syntactic priming in children) suggests that language selection may involve a shared process, but that selecting a specific linguistic structure may involve a separate process. This study provides somewhat fuzzy or indirect evidence. It is surprising that the authors did not discuss the many studies (see Paap et al. 2016 for a review) that correlate language switching with switching between two different nonverbal tasks. These consistently show that good language switchers are not necessarily good general task switchers. Relatedly, there is also a dissociation between aging and type of switching, as nonverbal switching can be severely impaired while language switching declines are very modest.

The underlying theory of bilingual advantages in EF is undermined if the competition between languages is not resolved by domain-general control mechanisms. A surprising development over the past few years is that the very existence of domain-general EF has been challenged, and this is not considered or addressed in this volume. This concern is discussed below, but we first need to consider the power of latent-variable analyses in identifying cognitive processes. Naomi P. Friedman’s Ch. 13 is a terrific introduction. One important topic is the problem of task impurity and how it can be mitigated with multiple measures of the same process yielding latent variables. There is a critical discussion of why large sample sizes are needed to detect even medium effect sizes because of task impurity and unreliability. The mechanics and consequences of the author’s unity and diversity model of EF are presented and include the devilish history of trying to extract a latent variable for inhibitory control. But does inhibitory control exist as a valid and reliable construct?

Rey-Mermet et al. (2018) urged us to ‘stop thinking about inhibition’. In a large-scale study they required each participant to complete six tasks assumed to reflect inhibition of prepotent responses and five assumed to reflect resistance to distraction. Model tests showed that the data are ambiguous as to: (a) whether there is one inhibition factor or two and (b) if there are two factors, whether they are correlated. Another problem pointed out by Rey-Mermet et al., both in their data and in other latent-variable analyses of the inhibition construct, is that for each latent variable, the loading for one task tends to dominate the others. Consequently, each latent variable represents mainly the variance of one task, and the remaining tasks are burdened with high error variances. From their view of this evidence, Rey-Mermet et al. suggested that nonverbal interference tasks do not measure a common underlying latent variable associated with general in-
hibitory control, but rather a highly task-specific ability to resolve the type of conflict instantiated in each task. Evidence supporting this position is presented in detail in Paap et al. 2021. To hammer home the point, even if the competition between languages is actively controlled by an inhibitory mechanism, it will not lead to far-transfer on nonverbal interference tasks if it is encapsulated within a bilingual language-control module. This reduces the phenomenon to the unremarkable claim that bilinguals get better at managing the competition between their two languages with practice. On the brighter side, Draheim et al. 2021 demonstrated substantial progress in developing a set of new EF tasks that cohere as a construct. Their primary strategy was to replace standard tasks using RT difference scores with adaptive versions that are more like unitary accuracy scores. They feel their data ‘demonstrate that attention control is a unitary concept’. This could put the hypothesis back into play.

Unicorns: the elusive necessary and sufficient conditions for generating advantages. One theme of the book is that the overall pattern of advantages and disadvantages appears inconsistent because only certain types of bilingual experience enhance domain-general EF and will affect different components of EF. For example, one section of Marton's chapter focuses on the possibility that bilingual advantages are contingent upon the level of proficiency attained in an L2: ‘Most studies compared bilingual and monolingual children’s outcomes in EF tasks, without considering differences in language proficiency among the bilingual speakers’ (266). Yet contrasting evidence suggests that the field is not so naive. For instance, measures of L1 and L2 are provided in almost every published study (albeit there are not always objective measures of both languages). Indeed, when Gunnerud et al. (2020) tested L2 proficiency as a moderator, their mediation analyses were based on 498 effect sizes! In contrast to the meta-analytic results, Marton reviews a far smaller set of results consistent with the conclusion that frequency of language use and age of acquisition moderate the effects of bilingualism on EF. In doing so she weaves together strands from different studies into a complex tapestry. Each significant language-group difference and each significant interaction is treated as a genuine piece of a jigsaw puzzle that reflects the real and complex relationships between facets of bilinguals and facets of EF. The problem is that these narratives rely on only a subset of a myriad of available studies that are replete with false positives and false negatives. The large-scale meta-analyses described earlier tested for many mediators and should have distilled any true sufficient conditions for generating bilingual advantages. They did not. For example, Gennerud et al. conclude that ‘none of the moderators that are theoretically believed to pinpoint the circumstances under which a bilingual advantage should occur were significantly related to differences in overall EF’ (2020:1075).

In what they call an ecological approach to bilingualism, Beatty-Martínez and Dussias propose that the consequences of bilinguals are driven by a combination of different control mechanisms that are differentially recruited by different aspects of bilingual experience. In this framework ‘it is not surprising that cognitive differences in bilinguals emerge under some circumstances but are absent in others’ (58). However, if any constellation of advantages and null results can be interpreted post hoc as support for the ecological approach, then this is nothing more than a recipe for more confirmation bias (Paap, Mason, et al. 2020). What we need is a solution to at least a part of the puzzle. That is, hypotheses must be advanced that bilingual experiences A, B, and C constitute a set of sufficient conditions for consistently producing advantages on specific measures (X, Y, and Z) of cognitive control that will replicate across laboratories.

Bak’s Ch. 6, ‘Cooking pasta in La Paz: Bilingualism, bias, and the replication crisis’, is wryly titled, artfully written, informative, five years old, and misleading in key places. Regarding one main theme, Bak asserts that ‘I question the assumption that different studies performed in different parts of the world should yield the same results. I argue that the environment in which an experiment is conducted can exert profound influence on its outcome’ (81). Bak is literally correct. But I fear it may induce us to tolerate inconsistent outcomes while making us feel that we understand the relationship between bilingualism and EF because it is complicated by the interaction of many factors. That is not sufficient. If we have an adequate theory of cooking pasta, then the theory must apply in all parts of the world. Indeed, cooking perfect pasta is a function of both cooking time and altitude. This is true everywhere. If bilingual advantages are real, then we should
eventually be able to state that they are a function of $X$, $Y$, and $Z$. Those necessary and sufficient conditions for generating bilingual advantages should apply anywhere.

Megan Zirnstein, Kinsey Bice, and Judith Kroll (Ch. 3) join the chorus of those suggesting that the consequences of bilingualism on cognitive control are responsive to the timing, proficiency, and pattern of use of individual bilinguals. In their abstract they optimistically promise that ‘the emerging pattern is complex but systematic, with the influence of language experience sometimes revealed in behavior but often seen only in brain activity’ (34). One emerging pattern described by Zirnstein et al. is that the consequences of bilingualism on cognition are determined not only by age of acquisition of L2 and/or L2 proficiency, but also by the frequency, context, and type of language switching. Ch. 3 presents many interesting examples but is hardly a systematic review. One informative discussion covers studies that measure both brain activity and behavior during different language-switching tasks. For example, studies by Blanco-Elorrieta and Pykkänen (2018) show that as the cues for language switching become more natural, the switch costs decrease. Blanco-Elorrieta and Pykkänen speculate that the only bilinguals who should show enhanced cognitive control are individuals who frequently find themselves in dual-language contexts, where they constantly need to respond to outside constraints that trigger frequent goal reconfigurations. But such conditions do not occur in natural conversations. Thus, it is not surprising that bilinguals who routinely switch languages in the same interactional context (dual-language bilinguals) do not, with one exception (Hartanto & Yang 2018), show advantages in nonverbal switch costs (Paap et al. 2016, Hartanto & Yang 2020, Lai & O’Brien 2020) compared to bilinguals who tend to use only one language in each interactional context.

In Ch. 9 Antonella Sorace examines a potential trade-off between inhibition and monitoring (i.e. the ability to adjust and refocus attention in a constantly changing environment). She suggests that early bilinguals develop in a way that is optimally suited to the use of two languages and the ability to switch between them, whereas late bilinguals develop enhanced inhibition because of the need to apply more inhibition to their dominant L1. Although Sorace reviews a study that fits this pattern, Paap, Johnson, and Sawi (2014) report a larger-scale study that does not (see also Paap et al. 2015). Paap et al. compared six language groups with median sample sizes of fifty-six: bilinguals with two native languages (i.e. quintessentially early), early bilinguals (before age six), late bilinguals (at or after age six), monolinguals with no L2 exposure, and two other groups of monolinguals with limited L2 exposure and proficiency. Measures typically assumed to reflect inhibitory control included antisaccade RT and accuracy, the flanker RT effect, and the Simon RT effect. The only measure of inhibitory control to show a significant main effect of group was the Simon effect, but the group with two native languages had the largest interference score, not the smallest. Furthermore, there were no differences in switch costs, and, consequently, it was not the case that the group with two native languages or the group of early bilinguals were better at switching than the late bilinguals or monolinguals. This is yet another example of a study that shows an interesting difference between two types of bilinguals, but the difference does not seem to matter in other labs and with larger samples.

Extracting signal from noise. Ellen Bialystok’s Ch. 2 argues that there is a strong signal that bilingual language control adapts domain-general control mechanisms if you know how to look for it. She effectively frames her arguments by analogy to Nate Silver’s (2012) explanation for why so many political predictions fail, namely, that you need to separate the signal (good data) from the noise (poor data). In this framework it follows that one does not want to make a prediction about an election outcome by simply counting the number of polls favoring candidate A versus candidate B. Rather, one should pay attention to the good polls and ignore the not-so-good polls. There are, however, a number of misapplications of this analogy. Bialystok chastises Paap, Johnson, and Sawi (2014) for ‘simply counting the number of studies that support each side’ (18). Bak points out the same fault and likewise cites Paap et al. 2014 as an example to be avoided. However, the point of the histograms reported by Paap et al. was not to simply show that far more published tests produced null results, but that the small number showing advantages were clustered in studies using small sample sizes. This is consistent with the fact that risky small sample sizes lead to more false positives when the null is true and is the opposite of what should
occur if there were real (but small) bilingual advantages in EF. In other words, Paap et al. (2014) were logically arguing that the studies showing positive effects tended to have small sample sizes and were likely acting as noise across the consistent signal of null results!

A similar canard appears in Ch. 5 by Whitford and Luk, who repeat Kroll and Bialystok’s (2013) claim that many of the null results obtained with between-groups comparisons may be due to violations of the assumption that the residuals are normally distributed. As Paap 2014 discusses in detail, Schmider et al. (2010) used computer simulations to examine the robustness of analysis of variance when the population distributions are markedly nonnormal. Across a range of effect sizes, the empirical type 1 error $\alpha$ and type 2 error $\beta$ remained constant under violations. The results show that Kroll and Bialystok were incorrect when they implied that nonnormality frequently leads to null effects when violated. Whitford and Luk are as well.

In her chapter’s conclusion Bialystok criticizes Paap and Greenberg (2013) for taking their null results as evidence that bilingualism has no effect on cognition, appealing to Fisher’s (1935) well-known point that ‘failure to reject the null hypothesis is not grounds for accepting the null hypothesis’ (37) (see also Bialystok 2016). The Basque Center on Cognition, Brain and Language (BCBL) and I have consistently used Bayes factor analyses in addition to null hypothesis testing. A Bayes factor is a ratio of the probability of the null hypothesis being true, given the data, over the probability of the alternative being true, given the data. While Bayes factors in the range of 0–3 are not very meaningful, going by Jeffrey’s (1961) guidelines, as they grow into the range of 3–10 they do provide substantial evidence that the null is true. In comparing bilinguals to monolinguals in Paap et al. 2014, we computed Bayes factors for twelve different measures of EF, and nine of the twelve produced Bayes factors in the range of 5–9. Bayes factors analyses are now routine in the reporting of ‘null’ effects.

Bialystok would also have us disregard the multitude of null results as noise, suggesting that they stem mostly from studies using young adults: ‘One simple possibility is that it is difficult for high functioning young adults to respond much faster than they already do—across most of these the mean RT for young adults is about 500 ms and it is difficult to see how an experiential difference could move an entire group to a significantly faster time’ (21). In a direct test of the ceiling hypothesis, Paap 2019 demonstrated that young adult university students were not at a performance ceiling.

According to Bialystok, another important step in separating the signal from the noise is selecting the right type of bilinguals. Factor-analytic methods were used to identify L2 proficiency and proportion of use as two important factors of bilingualism. Four groups of bilinguals were formed by a median split on both factors. These four groups of bilinguals were all compared to a monolingual group on a composite measure of EF, but only the group that was high on both L2 proficiency and balanced use was significantly better than the monolingual group. Bialystok implies that the bilinguals routinely tested by Paap and colleagues typically do not have both high proficiency and high use and hence null results should have been expected. In other words, the Paap studies are mostly noise. Even more specifically, Bialystok states that Paap et al. (2014) ‘reported that the bilinguals in their study used English about 72% of the time and had self-rated L2 proficiency of at least 4 out of 7 and the monolinguals used English about 90% of the time and had self-rated L2 proficiency of 2.5 out of 7’ (22). In the absence of clear background information, simple RT tasks do not discriminate between shades of bilingual experience.

A serious problem here is that Paap et al. (2014) did provide more information about their participants, but Bialystok’s summary is in part incomplete, in part inaccurate, and entirely misleading. Paap et al. 2014 was not a simple test of one group of bilinguals against a group of monolinguals. It involves a series of reanalyses of a database of 168 bilinguals and 216 monolinguals and twelve different measures of EF. In one highly relevant analysis, three groups of bilinguals and two groups of monolinguals were formed. In the high-proficiency group ($n = 120$), the mean L2 proficiency was 6.2 on a scale where 6 was native fluency and 7 was super fluency. The mean percentage of English spoken was 67.6%. Characterizing this group as having an L2 proficiency in the range of 4 to 7 (rather than as having a mean of 6.2) is very misleading. The ‘pure’ monolingual group ($n = 84$), with self-rated L2 proficiencies of 0 or 1 and a mean of 0.3, used English 99.8% of
the time. Across the twelve measures of EF there were never any significant differences between these two extreme groups. Furthermore, in another set of analyses Paap et al. treated bilingualism as a continuous variable. These regression analyses yielded nonsignificant coefficients for all dimensions of bilingualism (age of acquisition, number of languages, L2 proficiency, language balance, and percentage of use) across all of the measures of EF. Bialystok’s argument that the comparisons were limited to two ‘noisy’ groups and that there was an absence of clear background information is disingenuous.

Extensions to autism and aging. Aparna Nadig and Anna Maria Gonzalez-Berrero (Ch. 20) pull together and extend their recent published work on the relationship between bilingualism and autism spectrum disorder (ASD). Their most provocative finding is a bilingual advantage in the dimensional card-sort task with school-age children diagnosed with ASD, but no mirrored bilingual advantage in the typically developing control group or in the parental reports of deficits in EF in everyday life. This is a welcome outcome, as it not only provides confirmatory evidence that bilingualism does not exacerbate the problems associated with autism, but also suggests it may actually mitigate them. However, skeptics will hear echoes of Brysbaert’s (2020) recent plea for bilingual researchers to step up our game and recruit adequate sample sizes even when it is difficult and costly to do so. Although the authors are diligent in matching their groups, there were only ten participants in each group, and based on sample size alone it is not likely that the pattern of results across the four groups would replicate.

These results contrast with a study we recently published using autism traits to predict self-reported and objective measures of EF in a sample of 200 university students (Mason et al. 2020). Higher levels of autism traits were significantly associated with poorer EF scores, but only when EF was measured with self-reports. Furthermore, continuous measures of bilingualism (L2 proficiency and L2 use) were not associated with any of the five different objective measures of EF performance. The two studies differ in participant age (children versus college students), autism (diagnosed versus trait levels), self-report measures, and objective measures, but it is disconcerting that findings in the one study generally show opposite effects in the other.

Ch. 21, by Caitlin Wei-Ming Watson, Jennifer J. Manly, and Laura B. Zahodne, takes up the important question of whether bilingualism can delay the onset of cognitive decline and dementia. They place appropriate emphasis on the methodological divide that separates positive from negative findings, namely, that ‘[s]tudies that retrospectively identify dementia symptoms among patients in specialty memory clinics usually find significant differences between bilinguals and monolinguals whereas longitudinal studies that recruit participants from the community and prospectively report dementia incidence based on direct, comprehensive assessment do not’ (356). The several factors that likely contribute to this difference are carefully presented, and a section is deservedly allotted to the importance of actually measuring changes over time. It is worth adding that when Ramos et al. (2016) enticed a group of elderly (mean = 69 years) Spanish monolinguals to study and acquire Basque as an L2 for 5.5 hours per week for eight months, they showed no improvement on measures of EF from the pretest to the posttest.

I close with some observations on the chapter by Raymond M. Klein (Ch. 16), who poses two questions in his title: (i) What cognitive processes are likely to be exercised by bilingualism? His answers are inhibition, monitoring, and switching. (ii) Does this exercise lead to extralinguistic cognitive benefits? He finds ‘no consistent and convincing evidence’ (257). Klein’s answers are woven into an engaging historical narrative: from the revolutionary exposé by Peal and Lambert (1962), to the pioneering work chronicled by Ellen Bialystok (2001), to his involvement in the highly cited first report of bilingual advantages in adults, and on to his partnership with Matthew Hilchey in the first systematic reviews of the effects of bilingualism on monitoring and inhibition. The chapter closes with an appeal for understanding the considerable influence of the zeitgeist on the research hypotheses we formulate. Whatever our current zeitgeist, it is somewhat disconcerting that extremely productive and recognized leaders in cognitive psychology can look at the same landscape and see such different topologies. In any event, it is clearly a strength of this book that a more skeptical view is represented. That said, it would have benefited from even more balance.
REFERENCES


DUÑABEITIA, JON ANDONI; JUAN ANDRÉS HERNÁNDEZ; ENEKO ANTÓN; PEDRO MACIZO; ADELINA ESTÉVÉZ; LUIS J. FUENTES; and MANUEL CARREIRAS. 2014. The inhibitory advantage in bilingual children revisited: Myth or reality? Experimental Psychology 61.234–51. DOI: 10.1027/1618-3169/a000243.

DRECHSEL, CHRISTOPHER; JASON TSUKAHARA; JESSIE MARTIN; CODY MASHBURN; and RANDALL ENGLE. 2021. A toolbox approach to improving the measurement of attention control. Journal of Experimental Psychology: General 150(2).242–75. DOI: 10.1037/xge0000783.


PAAP, KENNETH R.; HUNTER A. JOHNSON; AND OLIVER Sawi. 2015. Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. Cortex 69.265–78. DOI: 10.1016/j.cortex.2015.04.014.


