



OPEN ACCESS

EDITED BY

Stephanie M. Carlson,
University of Minnesota Twin Cities,
United States

REVIEWED BY

Anthony Steven Dick,
Florida International University, United States
Scott R. Schroeder,
Hofstra University, United States
Sybren Spit,
University of Amsterdam, Netherlands

*CORRESPONDENCE

Samuel G. Marshall
✉ smarsh95@uwo.ca

†These authors have contributed equally to
this work

RECEIVED 15 July 2025

ACCEPTED 30 September 2025

PUBLISHED 24 October 2025

CITATION

Marshall SG and Bruce Morton J (2025) Back
to the test: Popper's neglected legacy in
bilingual advantage research.
Front. Dev. Psychol. 3:1666080.
doi: 10.3389/fdyps.2025.1666080

COPYRIGHT

© 2025 Marshall and Bruce Morton. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Back to the test: Popper's neglected legacy in bilingual advantage research

Samuel G. Marshall*† and J. Bruce Morton†

Cognitive Development and Neuroimaging Laboratory, University of Western Ontario, Department of Psychology, London, ON, Canada

Cognitive developmental science has made unprecedented progress in the last 50 years but has also seen many seminal findings fail to replicate. Adopting the bilingual advantage in children's attention control as a case study, we draw a connection between the replication crisis playing out in selected quarters of our field and a wavering commitment to the principles of classical hypothesis testing. Moving forward, we suggest open-science practices as a way of ensuring scientific hypotheses remain falsifiable and unconfirmed.

KEYWORDS

bilingual advantage, replication crisis, falsification, cognitive development, psychological development, hypothesis testing, Open Science

1 Introduction

Over the last 50 years, we have witnessed a breathtaking evolution in the empirical study of psychological development. Using an impressive array of behavioral, computational, comparative, genetic, epigenetic, and neurophysiological methods, developmental scientists have revealed: (1) sophisticated proficiencies in infants' perception of language, emotion, numeracy, and objects; (2) cognitive and neurophysiological mechanisms underlying the growth of higher-order cognitive abilities such as executive functioning and theory of mind; and (3) the sensitivity of development to variations in the quality of early experience (Golinkoff et al., 2013; Diamond, 2002; Wimmer and Perner, 1983; Tomasello et al., 2005; Xu and Tenenbaum, 2007; Dehaene-Lambertz et al., 2002; Saffran et al., 1996; Sorce et al., 1985; Zelazo and Carlson, 2012; Wynn, 1992; Gopnik and Meltzoff, 1997; Hackman et al., 2010).

At the same time, our field—like other branches of psychology—is facing a growing realization that many seminal findings cannot be reliably reproduced when studies are repeated using comparable methods. Indeed, according to one estimate, only 36–40% of seminal studies in psychology replicate successfully (Open Science Collaboration, 2015). To be sure, replication failures are not problematic in their own right—some replication failures should be expected when true underlying effect sizes are small—but can spawn crises if they are ignored. This is the situation we face in our field today.

2 The bilingual advantage in children's attention control: a replication crisis in developmental research

One striking illustration of the replication crisis in cognitive developmental research concerns the bilingual advantage in children's attention control. First reported over

50 years ago in the seminal work of [Peal and Lambert \(1962\)](#), evidence that bilingual children show an advantage in selective aspects of attention control relative to their monolingual counterparts gained widespread interest in 1999 with the publication of Ellen Bialystok's *Cognitive Complexity and Attentional Control in the Bilingual Mind* in the journal *Child Development*. The study compared Chinese-English bilingual and English monolingual kindergartners on a mental flexibility task called the Dimensional Change Card Sort. In this task, children begin by sorting cards one way and then are asked to switch and sort the same cards a different way. Consistent with the idea that bilingual language experience contributes to an advantage in children's attention control, bilingual children correctly switched more often than did monolingual children. This finding had a substantial impact on the study of bilingualism and attention development, as it suggested that the experience of controlling attention within the domain of language generalized to problems outside of language.

In the following years, evidence for the domain-general "effects" of bilingualism continued to accumulate, with bilinguals "outperforming" monolinguals on a variety of non-verbal attention tasks, including the Simon task, flanker tasks, ambiguous figures tasks, and saccadic eye movement tasks ([Bialystok et al., 2004](#); [Yang et al., 2011](#); [Wimmer and Marx, 2014](#)). Further study suggested these "effects" were not confined to childhood but could be observed in adults, the elderly, and preverbal infants ([Chung-Fat-Yim et al., 2017](#); [Schroeder and Marian, 2012](#); [Comishen et al., 2019](#)). Advocates for the bilingual advantage hypothesis trumpeted the "benefits" of bilingualism, emphasizing the power of bilingual experience to "reshape" neurocognitive function across the lifespan. And while the precise locus of these "effects" remained somewhat ill-defined, the consensus view was that bilingualism led to "improvements" in children's selective, or executive, attention—the very aspect of higher-order thinking bilingual children putatively engaged during daily language use.

Even within the golden age of bilingual advantage research, though, "confirmatory" evidence was often equivocal. In some studies, bilinguals were not only faster and more accurate than monolinguals on incongruent trials that require selective attention but also on congruent trials that do not require selective attention ([Bialystok et al., 2004](#); [Costa et al., 2008](#); for discussion, see [Hilchey and Klein, 2011](#)). In other studies, massive differences between bilinguals and monolinguals observed at the outset of testing disappeared in subsequent blocks ([Bialystok et al., 2004](#), Experiment 3). And still other studies found differences between bilingual and monolingual children but only after controlling for differences in L1 proficiency ([Carlson and Meltzoff, 2008](#)).

True uncertainty about the bilingual advantage hypothesis, however, surfaced with the publication of unapologetic null findings—studies that failed to detect any difference in bilingual and monolingual children's selective attention and reported the findings as such. The earliest of these was a small study that compared bilingual and monolingual children's performance on the Simon task ([Morton and Harper, 2007](#)). The authors acknowledged the importance of bilingual advantage research but noted that many studies risked confounding language status with a litany of unmeasured nuisance variables. Thus, the goal of the study

was to simply compare attentional control in monolingual and bilingual children after measuring and controlling for important confounds such as socio-economic status and immigration status. When carefully matched in this way, bilingual and monolingual children performed comparably on a version of the Simon task used in previous research ([Bialystok et al., 2004](#)). A few years later, [Paap and Greenberg \(2013\)](#) conducted a larger and more comprehensive test of the bilingual advantage hypothesis in adults. Across several experiments, adult bilinguals and monolinguals were administered non-verbal measures of attention control, including Simon, saccadic eye movement control, and flanker tasks. The authors then tested whether bilinguals showed faster and more accurate performance than monolinguals on trials that required attention control but comparable performance on all other trials—exactly the prediction made by the bilingual advantage hypothesis. Contrary to the bilingual advantage prediction, however, bilinguals and monolinguals performed comparably across all trial types. Then in 2019, [Dick et al. \(2019\)](#) used big data methods to test the bilingual advantage hypothesis in adolescents, drawing on a sample of 4,524 participants from the Adolescent Brain and Cognitive Development study. As expected, bilinguals had lower English (i.e., L1) vocabulary scores compared to monolinguals, but the groups were indistinguishable with respect to attention control.

Mounting uncertainty about the bilingual advantage hypothesis led some to revisit seminal studies and address lingering concerns about design and interpretation. Foremost was Bialystok's *Cognitive Complexity and Attentional Control in the Bilingual Mind* ([Bialystok, 1999](#)), which, as discussed earlier, had tested the bilingual advantage hypothesis by comparing attention control in English-Chinese bilingual and English monolingual kindergartners. The challenge in interpreting these findings is that beyond a difference in language status, English-Chinese bilingual and English monolingual kindergartners also differed in country of origin, a factor that predicts differences in kindergartners' attention control ([Sabbagh et al., 2006](#)). Thus, [Cho et al. \(2021\)](#) revisited [Bialystok \(1999\)](#) study by first comparing East Asian bilingual and Caucasian monolingual kindergartners on a measure of attention control and then extending the comparison to a sample of Korean monolinguals from South Korea. The results were clear. East Asian bilinguals outperformed Caucasian monolinguals in Canada, as was true in [Bialystok \(1999\)](#). However, East Asian bilinguals in Canada were indistinguishable from Korean monolinguals in South Korea. Thus, when language status and country of origin were unconfounded, the "bilingual" advantage reported by [Bialystok \(1999\)](#) appeared to be an East Asian advantage.

The most profound challenge to the bilingual advantage hypothesis, however, came in the form of several meta-analyses published in flagship journals of the APA and APS starting in 2018 ([Gunnerud et al., 2020](#); [Lehtonen et al., 2018](#); [Lowe et al., 2021](#)). Based on over 20 years of research on the bilingual advantage in adults and children, these studies all report that differences between bilinguals and monolinguals on measures of higher-order cognition, including attention control, are diminishingly small and subject to publication bias and methodological weaknesses. When published effect sizes were corrected for these influences, the effect of language status on higher-order cognition fell to zero. To be fair, these meta-analyses were extremely comprehensive and

included outcome measures that were never considered relevant by the bilingual advantage hypothesis—measures of, for example, planning and general executive functioning. At the same time, some authors did make a concerted effort to test for the bilingual advantage on measures that were clearly germane. Lowe et al. (2021), for example, tested for the bilingual advantage on measures of “executive attention”—a domain of higher-order cognition Bialystok (2017) argued was the central locus of language status effects—and found the overall effect of language status on these measures was indistinguishable from zero. Taken together, these meta-analytic findings should have, at a minimum, led to some reflection among advocates of the bilingual advantage hypothesis.

There is, however, no indication that advocates give proliferating null findings anything more than passing consideration. Beyond perfunctory reference to “challenges” (e.g., Koch et al., 2024) or “complexities” (see Festman et al., 2022), proof of the bilingual advantage remains a starting point in many current manuscripts. Reflecting on the last 30 years of bilingual advantage research, Bialystok (2025), for example, states that her main ideas have held up and there is now ample evidence confirming the lasting beneficial effects of bilingualism on selective attention. Alas, discourse on the bilingual advantage has reached an impasse. Advocates cling to ideas proposed over 30 years ago and claim that their empirical evidence confirms these ideas. Critics, meanwhile, question the credibility of the bilingual advantage hypothesis while calling attention to repeated replication failures.

As we look to the future of research in cognitive development, it is worth reflecting on the nature of this replication crisis. Is this intellectual conflict a natural part of an evolving science, something that should be embraced or even celebrated? Or does it reflect shortcomings in the way we currently practice our science? We submit the replication crisis surrounding the bilingual advantage in children’s attention is not indicative of a vibrant science at all but reflects serious weaknesses in the way science in this area has been conducted over the last 30 years. And the root of the problem driving this crisis of confidence is, in our view, unprincipled hypothesis testing.

3 The philosophical foundations of hypothesis testing

Hypothesis testing is a form of scientific reasoning developed by the philosopher Karl Popper in the early 20th century. Before then, scientific reasoning had been predominately inductive, moving from individual observations to general principles (Laudan, 1968; for more on induction, see Sprenger and Hartmann, 2019). Popper, however, like his predecessor David Hume, had reservations about inductive reasoning, because observational or confirmatory evidence alone provides no sound justification for general claims (Popper, 1959, 1962, 1976; Hume, 1739). Philosopher Bertrand Russell illustrated this with a story of a chicken that is fed every day. Using inductive reasoning, the chicken concludes that it will always be fed. This reasoning proves tragically wrong when the farmer eventually sacrifices the chicken to feed his family. Deductive reasoning, in contrast, would force the chicken to consider that unexpected events can happen, preventing it from being caught off guard. The chicken’s false confidence highlights

the limits of induction and motivated Popper to propose hypothesis testing as a more reliable foundation for scientific reasoning. In this framework, the value of a theory lies not in its capacity to produce confirming instances but instead in its ability to generate bold, testable predictions that can be refuted deductively through empirical observations. Hypothesis testing should challenge our assumptions and lead us to discard what is false, rather than accumulate confirmatory evidence.

However, in order for hypothesis testing to “work”—that is, to promote increasingly credible ideas—scientists need to avoid inductive inference and ensure that research hypotheses are genuinely falsifiable. Popper advocated for hypothesis testing precisely because it relies on deductive rather than inductive logic. Unlike inductive logic, deductive logic relies on truth-preserving structures such as *modus ponens*—affirming the antecedent—and *modus tollens*—denying the consequent. The epistemic strength of these logical forms derives from the fact that if the premises hold—or fail to hold—the conclusions necessarily follow. Consider a conditional statement, such as *if P then Q*, where *P* is the antecedent and *Q* is the consequent. If *P* is true, then *Q* is necessarily affirmed—*modus ponens*—whereas if *Q* is false, then *P* is necessarily denied—*modus tollens*. Inductive inference, by contrast, is not truth-preserving, as it is based on the reverse direction of inference—deriving the validity of a hypothesis (antecedent) from confirming observations (consequent) (Popper, 1959, 1962). For example, “If it rains, the road will be wet. The road is wet. Therefore, it must be raining.” This kind of argument—termed affirming the consequent—is inferential and fallacious—there are many reasons why the road might be wet beyond the weather. One strength of hypothesis testing, then, is that when practiced properly, it avoids inductive inference and relies on truth-preserving deductive logic.

Most importantly though, principled hypothesis testing requires that scientific propositions be falsifiable or susceptible to refutation. The proposition *if X, then Y* is falsifiable if, and only if, instances exist where *Y* does not follow *X* can be identified. Otherwise, the hypothesis is not falsifiable. This ensures that in building knowledge, we avoid affirming the consequent and adhere to the truth-preserving logic of *modus tollens*: if a theory generates a hypothesis, and that hypothesis fails, the consequent is denied, and the hypothesis is necessarily refuted. Without falsification, hypothesis testing yields little more than untested conjecture.

Beyond deductive logic and falsifiability, Popper argued that scientists also need to make bold predictions that *genuinely* risk falsification. As with any methodological instrument, falsification can be applied with varying degrees of rigor—at times sufficiently and other times superficially. This gradation, defined by Popper as a hypothesis’s degree of testability—or falsifiability—was based upon two key principles of a hypothesis: its empirical content—the degree to which it rules out possible outcomes—and its theoretical boldness—the degree to which it is open to refutation (Popper, 1959, 1962, 1983). While conceptually distinct, the roles of these principles are deeply intertwined. For example, hypotheses rich in empirical content are often inherently bold and face greater exposure to potential falsification (Vignero and Wenmackers, 2021). The important implication is that the falsifiability of a hypothesis is best understood as a continuum of epistemic significance that varies with the degree of its testability (Holtz and Monnerjahn, 2017).

At the top of this methodological continuum are hypotheses high in both content and boldness, which function as the predominant method for scientific progress. Consider, as an example, Jean Piaget's stage theory of cognitive development, which—while mostly falsified at this point—demonstrated high degrees of testability (Barrouillet, 2015). His theory of cognitive development proposed explicit hypotheses about the precise ages at which specific cognitive capacities would emerge. This boldness in its claims, mixed with the breadth of scenarios it excluded, exposed Piaget's hypothesis to a high degree of testability, which later resulted in major revisions to his underlying theory. From a Popperian perspective, this is the precise mechanism through which science is intended to operate. Theories are proposed with specific hypotheses; those hypotheses are tested, potentially falsified, and new theories are developed to replace those disconfirmed. The history of cognitive development has repeatedly illuminated how high degrees of testability foster this type of theoretical evolution, including in hypotheses about language acquisition, information processing, memory, and numerical cognition (Yang, 2004; Glenberg et al., 2013; Siegler et al., 2003).

Conversely, hypotheses at the lower end of this methodological continuum (i.e., low in both empirical content and boldness) have minimal degrees of testability and restrict scientific progress. Of these principles, Popper was particularly concerned about the effects of nominal theoretical boldness—or risk aversion. Without the risk of falsification, tests—even attempts at refutation—can do little to support the evolution of science (Popper, 1962). Risk aversion can emerge at two stages in a hypothesis's lifecycle: first, during its initial formulation, and later, in its resistance to disconfirming evidence. Some hypotheses at inception are risk averse—deliberately structured in ways that minimize their susceptibility to falsification. Take, for instance, pseudoscientific hypotheses such as astrology, which lack scientific value precisely because their predictions are so vague and flexible that they are inherently impossible to disconfirm. No amount of refutation can save these hypotheses from these inherent flaws. On the other hand, some hypotheses are formulated with scientific integrity—making bold, testable claims—but over time become risk-averse in response to falsifying evidence. Often, in these cases, advocates, upon evidence of falsification, attempt to preserve their hypotheses by reinterpreting them to introduce *ad hoc* assumptions and vagueness (i.e., reducing theoretical boldness), thereby undermining testability. Thus, these hypotheses, though once susceptible to refutation, become insulated from future falsification. As an example of one such hypothesis, Popper pointed to Karl Marx's theory of history. While first testable and falsifiable, following early attempts at its refutation, Marxists altered the theory to avoid future falsification (Popper, 1962). This conventionalist twist—as Popper defined it—may preserve the hypothesis, but only at the cost of its epistemic value and its support of scientific progress. Hypotheses at both these stages of their lifecycle expose an important truth about the principle of theoretical boldness within falsification: bold claims hold value, but only to the extent that scholars are willing to recognize and integrate disconfirming evidence. Without consistent application of these ideas, falsification becomes merely an empty formality, offering little more than the illusion of critical testing that stifles scientific progress

and allows long-falsified hypotheses to persist (LeBel et al., 2017).

Therefore, in Popperian terms, a healthy science necessitates principled hypothesis testing, which operates through two crucial principles—the avoidance of induction and the commitment to risky falsifiability. In the absence of these principles, scientific thinking risks stagnating.

4 The bilingual advantage and the absence of principled hypothesis testing

The replication crisis in bilingual advantage research illustrates how scientific thinking stagnates when hypothesis testing becomes unprincipled—that is, when researchers employ inductive—or confirmatory—logic (Rajtmajer et al., 2022; Earp and Trafimow, 2015) and propose non-falsifiable research hypotheses.

4.1 Language status differences do not confirm the “bilingual advantage”

Confirmatory logic is endemic in the bilingual advantage research literature. Consider Bialystok (2025) review of bilingual advantage findings. Referring to “modifications to cognitive performance in bilinguals” that “extend to nonverbal cognitive tasks,” she writes:

“There is now ample evidence both in behavioral and brain studies to support that claim. The notion that these effects would be continuous and tied to the extent of bilingual experience has been confirmed by neuroimaging studies showing linear correlation between modifications in brain structure and function and degree of bilingualism.”

Several aspects of this argument deserve critical examination. First, a correlation between two variables, such as brain structure and bilingualism, is no basis for inferring a causal association between those two variables. But beyond this, notice how evidence is granted confirmatory power in this passage. Correlational evidence does not corroborate the bilingual advantage hypothesis; it confirms it.

Perhaps more concerning, however, is that many studies of the bilingual advantage in children use inductive logic to support their conclusions. Consider the idea that bilinguals are advantaged in selective attention based on evidence that Chinese-English bilinguals outperform Caucasian monolinguals on the DCCS task (Bialystok, 1999)—this is a clear instance of affirming the consequent. It takes the following form:

If *P* (bilingual children are advantaged in attention control), then *Q* (bilingual children should achieve higher scores in selective attention tasks). *Q* is observed, therefore *P*.

But why, as Popper argues, is this logically fallacious?

The problem is that the Bialystok (1999) findings are consistent with the hypothesis that bilinguals are advantaged in attention control compared to monolinguals, but they are also consistent with the hypothesis that children of East Asian

origin are advantaged in attention control relative to Caucasian children from North America (Cho et al., 2021). Thus, the evidence is *consistent* with several hypotheses but does not *confirm* any one in particular. This is true not only of obviously confounded studies like Bialystok's seminal (Bialystok, 1999) study but even of carefully controlled experiments. Evidence is always consistent with several research hypotheses. At best, highly controlled experiments can *corroborate* a hypothesis, meaning that the hypothesis has survived previous falsification attempts and has gained greater credibility than other alternative explanations (Popper, 1959). Admittedly, disciplines vary in the extent to which their claims can be corroborated. In physics, for example, experiments can be methodologically very rigorous and generate results that corroborate a single hypothesis over others. Other disciplines, however, such as cognitive development and psychology, contend with greater conceptual uncertainty, larger measurement error, and smaller effect sizes, making it difficult to corroborate a single hypothesis over all others. Despite these differences, both psychology and physics operate within the same constraint: conclusions about hypotheses are only ever tentative—while some appear to approach the elusive “truth,” they are nonetheless *corroborated*, never confirmed.

For this reason, null hypothesis significance testing, which we putatively practice in the field of cognitive development, tests the null hypothesis that monolinguals and bilinguals are indistinguishable in performance, not the alternative or research hypothesis that bilinguals are advantaged in attention control. We proceed by rejecting ideas that are demonstrably false, not by confirming ideas we believe are true.

4.2 The bilingual advantage is not a falsifiable hypothesis

A second reason why bilingual advantage research is beset by a replication crisis is that the bilingual advantage hypothesis is not falsifiable. To be sure, hypotheses are not categorically falsifiable or unfalsifiable, and there have been some genuine attempts to falsify the bilingual advantage hypothesis (for example, Yang et al., 2011). However, the hypothesis has broadened and narrowed since the 1990s. It initially focused on “attention control” (Bialystok, 1992), broadened to encompass executive functioning generally (Bialystok, 1999), and then focused again on various aspects of attention, including “selective attention” (Bialystok, 2025), “executive attention” (Bialystok, 2017), and “attentional disengagement” (Grundy et al., 2017). The hypothesis has also migrated away from differences between bilinguals and monolinguals toward continuous differences between bilinguals with varying degrees of second language experience (Surrain and Luk, 2019). One could argue that these sorts of changes to the bilingual advantage hypothesis illustrate both its falsifiability and theoretical development.

This is true, however, only in the most superficial sense, as modifications to the hypothesis have been introduced primarily to discount the relevance of replication failures (for discussion, see Bialystok, 2025). For example, citing evidence that bilingual and

monolingual adults perform comparably on executive functioning tasks, Bialystok (2025, 2017) claimed that the bilingual advantage is not related to executive functioning but to selective or executive attention. The problem with this argument is that studies that failed to replicate the bilingual advantage were portrayed as studies of executive functioning (Paap and Greenberg, 2013) but had, in fact, used tasks (e.g., Simon, Flanker, and switching) considered elsewhere to be operationalizations of “selective” or “executive attention” (Bialystok, 2017). In fact, in our own meta-analytic review of the bilingual advantage literature, we adopted operational definitions of “executive attention” published by Bialystok (2017) and compared language status effects in this domain with language status effects in executive functioning domains (Lowe et al., 2021). The distinction proved to be uninformative. Language status effects within executive attention and executive function domains were both indistinguishable from zero after correcting for publication bias and study quality. If null findings from such a test do not falsify the hypothesis, it is not clear what evidence could. From our vantage point then, there is simply not sufficient clarity about what executive attention is and how it should be measured to allow independent researchers to conceive an observation that would falsify the bilingual advantage hypothesis. As such, the bilingual advantage in children's attention control is, in our opinion, not a falsifiable hypothesis.

5 Restoring credibility to bilingual advantage research through principled hypothesis testing and Open Science

So far, we have painted a somewhat grim picture of the state of bilingual advantage research. The field is awash in contradictory evidence and deadlocked in a tired debate about ideas that, on some accounts (Bialystok, 2025) have not changed appreciably in the last 30 years. This not only reflects negatively on our field as a science but also has real-world implications. Beliefs in the transformative neurocognitive “effects” of bilingualism permeate popular culture and influence the decisions of parents and policymakers around the world (Mehmedbegovic, 2018; Goldenberg and Wagner, 2015; Woll and Wei, 2019). This is not a legacy we should take comfort in.

At the same time, restoring vitality and credibility to this research area is not beyond reach and may be as straightforward as adhering to more principled hypothesis testing. We offer five directives for achieving this goal moving forward.

5.1 The bilingual advantage hypothesis needs to be falsifiable

First and foremost, the bilingual advantage hypothesis needs to be stated in falsifiable terms, starting with well-defined concepts at a general level and extending to detailed predictions in the context of specific experiments.

With regard to concepts, we need clear definitions of key independent and dependent variables, including, at a minimum, bilingualism and attention control. To be sure, definitions of these terms do exist, but these definitions can, and often do,

admit uncertainty. Imprecise definitions undermine both the empirical content and the boldness of hypotheses by decreasing the number of outcomes a hypothesis excludes and the likelihood a hypothesis will be proven wrong. Consider the concept of bilingualism. In recent years, researchers have moved toward a definition that emphasizes continuous inter-individual variation in the ability to understand and use a “second language” (Surrain and Luk, 2019). Such a definition certainly makes sense in view of the heterogeneity among individuals with second language proficiency. But where does this leave monolinguals traditionally defined? If second language proficiency is truly continuous, is there such a thing as a monolingual speaker? Don’t most English speakers understand and use “*Gesundheit*” as a response to a sneeze or experience “*Shadenfreude*” when an adversary suffers an unwelcome setback? If every speaker has at least some second language proficiency, does it even make sense to distinguish between monolinguals and bilinguals? And what is a “second language”? Speakers, for example, routinely select words in the interest of social appropriateness, using some words in some contexts (while socializing with friends) but not others (while at church). Parents routinely switch to a different form of communication when talking with infants and then switch back to conventional language when talking with older children and adults. And many speakers (for example, in China/Germany) use a general language form in public (*Mandarin/Hochdeutsch*) but a dialect (*Shanghaiese/Plattdeutsch*) in private. Are these examples of second languages? Similar challenges exist for the concept of attention. The bilingual advantage hypothesis has narrowed from executive function to “attention” in recent years, but there are still a wide variety of terms used to conceptualize attention, including “selective attention,” “attentional disengagement,” and “executive attention.” It remains unclear whether these are different forms of attention or all facets of one larger function. In the absence of clear conceptual definitions, the bilingual advantage hypothesis is difficult to test and near impossible to falsify.

Beyond conceptual definitions, there is also a need for clarity in the operational definitions of bilingualism, attention, and the bilingual advantage itself. If second language proficiency is continuous, and all speakers show at least some understanding and use of a second language, how should researchers sample the population? Should we sample and then test for continuous variation only among bilinguals (e.g., Oh et al., 2023), or should every speaker be included in a sample given that the distinction between bilinguals and monolinguals is arbitrary? What tasks should be used to operationalize attention? Do all trials matter, or do some trials—for example, incongruent trials—matter more than others—for example, congruent trials (for discussion, see Hilchey and Klein, 2011)? Finally, what pattern of performance should be observed? Should more proficient bilinguals (or just bilinguals) be advantaged in attention control compared to less proficient bilinguals (or just monolinguals)? Should researchers predict a main effect of second language proficiency or a higher-order interaction between second language proficiency and trial type? And if the latter, what higher-order interaction specifically? Enhancing the operational definitions associated with the bilingual advantage hypothesis is crucial for increasing its degree of

testability and thereby its falsifiability. More precise operational terms could exclude more outcomes (increasing empirical content) and increase the hypothesis’s exposure to refutation (increasing theoretical boldness), sharpening what counts as confirming—and disconfirming—evidence and heightening its degrees of testability.

Providing clear definitions of constructs like attention and bilingualism is beyond the scope of the current manuscript—indeed attempts to define these constructs have a long history in psychology. Instead, we highlight promising parts of this larger narrative that can hopefully serve as a starting point for an overdue discussion (Wagner et al., 2022; Sanches de Oliveira and Bullock Oliveira, 2022). Let’s begin with executive attention. This is a term for which there are actually relatively clear conceptual and operational definitions, and that figures prominently in discussions of the bilingual advantage (Bialystok, 2017). Executive attention as a concept was first put forward by Randy Engle in the context of discussions of working memory capacity (Engle, 2002) and refers to a domain-general ability to maintain task goals and suppress distractions or conflicting responses amid interference. Operational definitions encompass different facets of the general concept and include measures of: (1) interference control (e.g., Stroop, flanker, and sorting (or rule-use) tasks); (2) goal maintenance under distraction (e.g., complex span and continuous performance tasks); and (3) response inhibition (e.g., stop-signal reaction time and anti-saccade tasks). If, as hypothesized, the bilingual advantage is confined to the domain of executive attention (Bialystok, 2025, 2017), then associations between language status and attention should be confined to these domains and tasks.

With regards to defining bilingual language status, we suggest abandoning the distinction between bilinguals (or multilinguals) and monolinguals and instead adopting Surrian and Luk’s definition of multilingualism as continuous inter-individual variation in the ability to understand and use a “second language” (Surrain and Luk, 2019). Given this definition, “monolinguals” traditionally defined should not be excluded from studies of multilingualism (e.g., Oh et al., 2023) but simply assigned to the low-end of the multilingual continuum (e.g., Grundy et al., 2017). Admittedly, this definition is blind to the many ways in which individuals at the high-end of the continuum could differ from each other in terms of their daily language experience. Characterizing and testing the relevance of the multidimensional nature of multilingualism for attention control is therefore an important challenge for future research. Finally, contrary to arguments that tests of the bilingual advantage should focus on continuous variation in multilingualism, we submit that a comparison of extreme groups remains the most sensitive design for detecting hypothesized associations between language experience and attention control. The only circumstance in which this would not be true would be one in which the association between language experience and attention control was non-linear, with individuals with moderate L2-proficiency having higher attention scores than both individuals with the lowest and the highest L2-proficiency—in short, an inverted U-shaped curve. This seems highly implausible.

Greater clarity in the conceptual and operational definitions of executive attention and bilingualism—will lend greater testability and falsifiability to the bilingual advantage hypothesis.

5.2 Confirmatory logic in the interpretation of evidence should be curtailed

Second, use of confirmatory logic in the interpretation of bilingual advantage evidence needs to be scaled back considerably. Higher attention scores for bilinguals relative to monolinguals do not confirm the bilingual advantage, nor do associations between continuous measures of second language proficiency and attention. Evidence of this kind—to the extent it can be replicated—is certainly consistent with these hypotheses but is by no means confirmatory. Insisting that evidence confirms hypotheses simply undermines the credibility of our science. There is also a need for greater caution regarding the kinds of conclusions drawn from available evidence. The bilingual advantage literature is filled with references to “the beneficial effects” of multilingualism on the developing mind and brain, and so on. Indeed, the very term “bilingual advantage” implies that bilingualism is linked to improvements rather than mere differences. Beyond the fact that science is not a moral enterprise that arbitrates between “good” and “bad,” between-group and individual differences designs provide no basis for inferring, let alone confirming, causal associations.

Therefore, in our view, there needs to be a much more judicious use of terminology in the discussion of bilingualism research. Between-group comparisons and correlational designs that form the balance of bilingual advantage research can challenge the null hypothesis (i.e., that language experience and cognition are unrelated) but do not confirm any alternatives. This is especially true of hypotheses concerning causal associations between bilingualism and benefits in cognition. Between-group comparisons and correlational designs generate evidence of associations, not beneficial effects. Moving forward, authors should discuss correlational evidence as consistent with hypothesized associations, not as a confirmation of beneficial effects.

5.3 Hypotheses, analyses, and predictions should be pre-registered

A third corrective that should be adopted by bilingual advantage researchers is the use of pre-registration through the Open Science Framework (<https://osf.io/>). Pre-registration allows researchers to publicly post the details of a study, including hypotheses, conceptual and operational definitions, sampling procedures, analyses, and predictions, prior to data collection. By committing to the details of their studies in advance, researchers promote greater transparency with respect to their methods and conclusions.

Ideally, pre-registration of bilingual advantage research should take the form of registered reports (Chambers, 2013). Registered reports are a form of publication developed in response to the replication crisis, in which authors report on their studies in two stages. In the first stage, or Stage 1 Registered Report (RR), authors report the research question, hypotheses, sampling procedures, analyses, and predictions they will follow in their study. The report is peer-reviewed and revised before publication. Data is then collected and analyzed according to the procedures described in

the Stage 1 RR, and the results are written up as a Stage 2 RR. Importantly, publication of the Stage 2 RR is not contingent on whether the results are consistent with what was predicted in the Stage 1 RR.

Adopting RRs as a standard for the publication of bilingual advantage research would lift research standards in the area considerably. First and foremost, by requiring the details of study designs and data analyses to be declared in advance of data acquisition, RRs would help to ensure hypotheses concerning the bilingual advantage were stated in falsifiable terms. Second, RRs would help to facilitate replication studies. Stage 1 RRs require methodological procedures to be reported in sufficient detail to allow other researchers to undertake the described study. Finally, by ensuring findings were published regardless of the outcomes, RRs would help mitigate publication bias. Through these means, pre-registration can help to facilitate principled hypothesis testing, thereby ensuring that researchers adhere to well-established Popperian standards.

Pre-registration of a study in the form of an RR is admittedly time-consuming, and not all peer-reviewed journals accept RRs. This can present challenges, especially for early-career researchers who need to demonstrate immediate productivity. As a result, changes to graduate program requirements, tenure-evaluation processes, and journal policies may need to be instituted before RRs are more widely adopted as a framework for developing and communicating research ideas.

5.4 Direct replication studies should be undertaken

As a fourth corrective, more direct replication studies are required. Direct replication studies are studies that test whether previously reported findings can be reproduced when a study is conducted again using identical methods (Lindsay, 2017; Derksen and Morawski, 2022). Given that language status effects on cognition are diminishingly small (Gunnerud et al., 2020; Lehtonen et al., 2018; Lowe et al., 2021), and psychological measurement in young children can be imprecise, effect size estimates generated by any particular research design will vary considerably from study to study. Therefore, it may be more profitable to generate a distribution for the effect size associated with a specific design through repeated replication than tying interpretations to any single experimental outcome. Given that there are now journals dedicated to publishing registered replication reports (Lindsay, 2017), this may be a productive avenue to pursue moving forward. At a minimum, it would ensure that methods and analytical procedures were reported in sufficient detail to permit repeated replication across different sites.

5.5 Greater use of Bayesian hypothesis testing may be advantageous

As a fifth and final corrective, advocates of the bilingual advantage hypothesis should make greater use of Bayesian

hypothesis testing. Critics of the bilingual advantage hypothesis routinely call attention to studies that find no difference between bilingual and monolingual children as evidence that the bilingual advantage hypothesis is false. When studies are undertaken with a frequentist null hypothesis significance testing framework, the absence of group differences provides no evidence in support of the null, since the null hypothesis in this framework is always assumed to be true. As such, the absence of group differences in studies of the bilingual advantage does little to challenge the bilingual advantage hypothesis. Bayesian hypothesis testing, by contrast, allocates credibility to a range of hypotheses, including the null hypothesis, prior to data collection (Fornacon-Wood et al., 2022), and then re-allocates credibility among these hypotheses in light of the data. As such, the credibility of the null hypothesis can and will increase given an absence of a group differences.

How well Bayesian statistics aligns with Popperian principles of hypothesis testing, however, remains unclear. Updating beliefs in the face of data places Bayesian hypothesis testing firmly within the realm of inductive inference (Hayes et al., 2010). Additionally, the issue of falsifiability within Bayesian methods is far from settled. For example, what constitutes a risky, falsifiable prior, and how can the posterior reflect the survival of such predictions given its inductive foundations? Thus, Bayesian inference should be considered an important complement to conventional null hypothesis significance testing but should not be naively embraced as a statistical panacea.

6 Toward a healthier science: restoring hypothesis testing in cognitive development

The five correctives outlined above mark what, in our view, are important steps for strengthening bilingual advantage research. Collectively, they will help ensure studies in this field—and perhaps the study of cognitive development more generally—observe the principles of hypothesis testing set out by Karl Popper. We submit that all scholars—both advocates and critics of the bilingual advantage hypothesis—should consider adopting at least some of these scientific practices moving forward. Doing so will, we believe, help restore vitality and credibility to a field deadlocked in controversy.

As a final note, we have focused on the case of the bilingual advantage in our analysis of the replication crisis confronting the field of cognitive development. While other research areas within the field are also grappling with problems of replicability (for discussion, see Baillargeon et al., 2018; Kulke et al., 2018; Watts et al., 2018), it is, of course, unclear whether our insights extend to or are even relevant to these other areas. We leave that for the reader to decide.

References

Baillargeon, R., Buttelmann, D., and Southgate, V. (2018). Invited commentary: interpreting failed replications of early false-belief findings: methodological and

7 Conclusion

Cognitive development has made substantial contributions to the understanding of human cognition. However, like other areas of research, our field must come to terms with the fact that many seminal findings do not replicate. Looking into the future, we encourage all scholars to embrace principled hypothesis testing—where scholars reject inductive practices and embrace risky falsification—as the preeminent method for navigating the current replication crisis.

Author contributions

SM: Conceptualization, Writing – original draft, Writing – review & editing. JBM: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the NSERC Discovery Grant awarded to JBM.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

theoretical considerations. *Cogn. Dev.* 46, 112–124. doi: 10.1016/j.cogdev.2018.06.001

- Barrouillet, P. (2015). Theories of cognitive development: from Piaget to today. *Dev. Rev.* 38, 1–12. doi: 10.1016/j.dr.2015.07.004
- Bialystok, E. (1992). “Selective attention in cognitive processing: the bilingual edge,” in *Advances in Psychology. North-Holland (Cognitive Processing in Bilinguals)*, ed. R. J. Harris (Amsterdam: North-Holland), 501–513. doi: 10.1016/S0166-4115(08)61513-7
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Dev.* 70, 636–644. doi: 10.1111/1467-8624.00046
- Bialystok, E. (2017). The bilingual adaptation: how minds accommodate experience. *Psychol. Bull.* 143, 233–262. doi: 10.1037/bul0000099
- Bialystok, E. (2025). Selective attention in cognitive processing revisited: a long-term re-evaluation. *Brain Lang.* 261:105518. doi: 10.1016/j.bandl.2024.105518
- Bialystok, E., Craik, F. I., Klein, R., and Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the simon task. *Psychol. Aging* 19, 290–303. doi: 10.1037/0882-7974.19.2.290
- Carlson, S. M., and Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Dev. Sci.* 11, 282–298. doi: 10.1111/j.1467-7687.2008.00675.x
- Chambers, C. D. (2013). Registered reports: a new publishing initiative at cortex. *Cortex* 49, 609–610. doi: 10.1016/j.cortex.2012.12.016
- Cho, I., Park, J., Song, H. J., and Morton, J. B. (2021). Disentangling language status and country-of-origin explanations of the bilingual advantage in preschoolers. *J. Exp. Child Psychol.* 212:105235. doi: 10.1016/j.jecp.2021.105235
- Chung-Fat-Yim, A., Sorge, G. B., and Bialystok, E. (2017). The relationship between bilingualism and selective attention in young adults: evidence from an ambiguous figures task. *Quart. J. Exp. Psychol.* 70, 366–372. doi: 10.1080/17470218.2016.1221435
- Comishen, K. J., Bialystok, E., and Adler, S. A. (2019). The impact of bilingual environments on selective attention in infancy. *Dev. Sc.* 22:e12797. doi: 10.1111/desc.12797
- Costa, A., Hernández, M., and Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition* 106, 59–86. doi: 10.1016/j.cognition.2006.12.013
- Dehaene-Lambertz, G., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015. doi: 10.1126/science.1077066
- Derksen, M., and Morawski, J. (2022). Kinds of replication: examining the meanings of ‘conceptual replication’ and ‘direct replication.’ *Perspect. Psychol. Sci.* 17, 1490–1505. doi: 10.1177/17456916211041116
- Diamond, A. (2002). “Normal development of prefrontal cortex from birth to young adulthood: cognitive functions, anatomy, and biochemistry,” in *Principles of Frontal Lobe Function*, eds. D. T. Stuss and R. T. Knight (Oxford: Oxford University Press), 466–503. doi: 10.1093/acprof:oso/9780195134971.003.0029
- Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., et al. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nat. Hum. Behav.* 3, 692–701. doi: 10.1038/s41562-019-0609-3
- Earp, B. D., and Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Front. Psychol.* 6:621. doi: 10.3389/fpsyg.2015.00621
- Engle, R. W. (2002). Working memory capacity as executive attention. *Curr. Direct. Psychol. Sci.* 11, 19–23. doi: 10.1111/1467-8721.00160
- Festman, J., Czapka, S., and Winsler, A. (2022). How many moderators does it take till we know... that too many bilingual advantage effects have died?” in *Understanding Variability in Second Language Acquisition, Bilingualism, and Cognition* (New York, NY: Routledge), 80–127. doi: 10.4324/9781003155683-4
- Fornacon-Wood, I., Mistry, H., Johnson-Hart, C., Faivre-Finn, C., O’Connor, J. P. B., and Price, G. J. (2022). Understanding the differences between bayesian and frequentist statistics. *Int. J. Radiat. Oncol. Biol. Phys.* 112, 1076–1082. doi: 10.1016/j.ijrobp.2021.12.011
- Glenberg, A. M., Witt, J. K., and Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspect. Psychol. Sci.* 8, 573–585. doi: 10.1177/1745691613498098
- Goldenberg, C., and Wagner, K. (2015). Bilingual education: reviving an American tradition. *Am. Educ.* 39:28.
- Golinkoff, R. M., Ma, W., Song, L., and Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: what have we learned?. *Perspect. Psychol. Sci.* 8, 316–339. doi: 10.1177/1745691613484936
- Gopnik, A., and Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Grundy, J. G., Anderson, J. A. E., and Bialystok, E. (2017). Neural correlates of cognitive processing in monolinguals and bilinguals. *Ann. N. Y. Acad. Sci.* 1396, 183–201. doi: 10.1111/nyas.13333
- Gunnerud, H. L., Ten Braak, D., Reikerås, E. K. L., Donolato, E., and Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychol. Bull.* 146, 1059–1083. doi: 10.1037/bul0000301
- Hackman, D. A., Farah, M. J., and Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nat. Rev. Neurosci.* 11, 651–659. doi: 10.1038/nrn2897
- Hayes, B. K., Heit, E., and Swendsen, H. (2010). Inductive reasoning. *WIREs Cogn. Sci.* 1, 278–292. doi: 10.1002/wcs.44
- Hilchey, M. D., and Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychon. Bull. Rev.* 18, 625–658. doi: 10.3758/s13423-011-0116-7
- Holtz, P., and Monnerjahn, P. (2017). Falsificationism is not just “potential” falsifiability, but requires “actual” falsification: social psychology, critical rationalism, and progress in science. *J. Theory Soc. Behav.* 47, 348–362. doi: 10.1111/jtsb.12134
- Hume, D. (1739). *A Treatise of Human Nature (1739–40)* (Oxford: Oxford University Press). doi: 10.1093/oseo/instance.00046221
- Koch, M. J., Kersten, K., and Greve, W. (2024). An emotional advantage of multilingualism. *Bilingualism: Lang. Cogn.* 27, 950–963. doi: 10.1017/S1366728923000937
- Kulke, L., Reiß, M., Krist, H., and Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cogn. Dev.* 46, 97–111. doi: 10.1016/j.cogdev.2017.09.001
- Laudan, L. (1968). Theories of scientific method from plato to mach: a bibliographical review. *Hist. Sci.* 7, 1–63. doi: 10.1177/007327536800700101
- LeBel, E. P., Berger, D., Campbell, L., and Loving, T. J. (2017). Falsifiability is not optional. *J. Pers. Soc. Psychol.* 113, 254–261. doi: 10.1037/pspi0000106
- 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. *Psychol. Bull.* 144, 394–425. doi: 10.1037/bul0000142
- Lindsay, D. S. (2017). Preregistered direct replications in psychological science. *Psychol. Sci.* 28, 1191–1192. doi: 10.1177/0956797617718802
- Lowe, C. J., Cho, I., Goldsmith, S. F., and Morton, J. B. (2021). The bilingual advantage in children’s executive functioning is not related to language status: a meta-analytic review. *Psychol. Sci.* 32, 1115–1146. doi: 10.1177/0956797621993108
- Mehmedbegovic, D. (2018). “What every policymaker needs to know about the cognitive benefits of bilingualism,” in *Languages after Brexit: How the UK Speaks to the World*, ed. M. Kelly (Cham: Springer International Publishing), 109–124. doi: 10.1007/978-3-319-65169-9_10
- Morton, J. B., and Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Dev. Sci.* 10, 719–726. doi: 10.1111/j.1467-7687.2007.00623.x
- Oh, J. H. J., Bertone, A., and Luk, G. (2023). Reading comprehension and cognitive correlates in multilingual children and adolescents. *Transl. Issues Psychol. Sci.* 9, 397–408. doi: 10.1037/tps0000375
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- Paap, K. R., and Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cogn. Psychol.* 66, 232–258. doi: 10.1016/j.cogpsych.2012.12.002
- Peal, E., and Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychol. Monogr.: Gen. Appl.* 76, 1–23. doi: 10.1037/h0093840
- Popper, K. R. (1959). *The Logic of Scientific Discovery*, ed. H. P. Group. New York, NY: Routledge. doi: 10.1063/1.3060577
- Popper, K. R. (1962). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.
- Popper, K. R. (1976). *Unended Quest: An Intellectual Autobiography*. New York, NY: Routledge.
- Popper, K. R. (1983). *Realism and the aim of science*. Totowa, NJ: Rowman and Littlefield.
- Rajtmajer, S. M., Errington, T. M., and Hillary, F. G. (2022). How failure to falsify in high-volume science contributes to the replication crisis. *eLife* 11:e78830. doi: 10.7554/eLife.78830.sa2
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., and Lee, K. (2006). The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychol. Sci.* 17, 74–81. doi: 10.1111/j.1467-9280.2005.01667.x
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926
- Sanches de Oliveira, G., and Bullock Oliveira, M. (2022). Bilingualism is always cognitively advantageous, but this doesn’t mean what you think it means. *Front. Psychol.* 13:867166. doi: 10.3389/fpsyg.2022.867166
- Schroeder, S. R., and Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *J. Cogn. Psychol.* 24, 591–601. doi: 10.1080/20445911.2012.669367
- Siegler, R., DeLoache, J., and Eisenberg, N. (2003). *How Children Develop*. New York, NY: Worth Publishers.

- Sorce, J. F., Emde, R. N., Campos, J. J., and Klinnert, M. D. (1985). Maternal emotional signaling: its effect on the visual cliff behavior of 1-year-olds. *Dev. Psychol.* 21, 195–200. doi: 10.1037/0012-1649.21.1.195
- Sprenger, J., and Hartmann, S. (2019). “Confirmation and induction,” in *Bayesian Philosophy of Science*, eds. J. Sprenger and S. Hartmann (Oxford: Oxford University Press), 41–66. doi: 10.1093/oso/9780199672110.003.0001
- Surrain, S., and Luk, G. (2019). Describing bilinguals: a systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Lang. Cogn.* 22, 401–415. doi: 10.1017/S1366728917000682
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; Discussion 691–735. doi: 10.1017/S0140525X05000129
- Vignero, L., and Wenmackers, S. (2021). Degrees of riskiness, falsifiability, and truthlikeness: a neo-Popperian account applicable to probabilistic theories. *Synthese* 199, 11729–11764. doi: 10.1007/s11229-021-03310-5
- Wagner, D., Bialystok, E., and Grundy, J. G. (2022). What is a language? Who is bilingual? Perceptions underlying self-assessment in studies of bilingualism. *Front. Psychol.* 13:863991. doi: 10.3389/fpsyg.2022.863991
- Watts, T. W., Duncan, G. J., and Quan, H. (2018). Revisiting the marshmallow test: a conceptual replication investigating links between early delay of gratification and later outcomes. *Psychol. Sci.* 29, 1159–1177. doi: 10.1177/0956797618761661
- Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5
- Wimmer, M. C., and Marx, C. (2014). Inhibitory processes in visual perception: a bilingual advantage. *J. Exp. Child Psychol.* 126, 412–419. doi: 10.1016/j.jecp.2014.03.004
- Woll, B., and Wei, L. (2019). *Cognitive Benefits of Language Learning: Broadening Our Perspectives*. The British Academy, 1–87. Available online at: <https://www.thebritishacademy.ac.uk/documents/287/Cognitive-Benefits-Language-Learning-Final-Report.pdf> (Accessed August 6, 2025).
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature* 358, 749–750. doi: 10.1038/358749a0
- Xu, F., and Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychol. Rev.* 114, 245–272. doi: 10.1037/0033-295X.114.2.245
- Yang, C. D. (2004). Universal Grammar, statistics or both?. *Trends Cogn. Sci.* 8, 451–456. doi: 10.1016/j.tics.2004.08.006
- Yang, S., Yang, H. J., and Lust, B. (2011). Early childhood bilingualism leads to advances in executive attention: dissociating culture and language. *Bilingualism: Lang. Cogn.* 14, 412–422. doi: 10.1017/S1366728910000611
- Zelazo, P. D., and Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: development and plasticity. *Child Dev. Perspect.* 6, 354–360. doi: 10.1111/j.1750-8606.2012.00246.x