Monolingual comparative normativity in bilingualism research is out of *"control"*: Arguments and alternatives

Jason Rothman[1,2], Fatih Bayram[1], Vincent DeLuca[1], Grazia Di Pisa[3], Jon Andoni Duñabeitia[2,1], Khadij Gharibi[1], Jiuzhou Hao[1], Nadine Kolb[1,4], Maki Kubota[1], Tanja Kupisch[3,1], Tim Laméris[5], Alicia Luque[2], Brechje van Osch[1], Sergio Miguel Pereira Soares[6], Yanina Prystauka[1], Deniz Tat[7], Aleksandra Tomić[1], Toms Voits[1], Stefanie Wulff[8,1]

UiT The Arctic University of Norway[1], Universidad Nebrija[2], University of Konstanz[3], , University of Stavanger[4], University of Cambridge[5], Max Planck Institute for Psycholinguistics[6], Leiden University[7], University of Florida[8]

Abstract

Herein, we contextualize, problematize and offer some insights for moving beyond the problem of monolingual comparative normativity in (psycho)linguistic research on bilingualism. We argue that, in the vast majority of cases, juxtaposing (functional) monolinguals to bilinguals fails to offer what the comparison is supposedly intended to do: meet the standards of empirical control in line with the scientific method. Instead, the default nature of monolingual comparative normativity has historically contributed to inequalities in many facets of bilingualism research and continues to impede progress on multiple levels. Beyond framing our views on the matter, we offer some epistemological considerations and methodological alternatives to this standard practice that improve empirical rigor while fostering increased diversity, inclusivity and equity in our field.

Introduction

As is the case for all papers, this one is written from the perspective of its authors, which inevitably entails biases stemming from our backgrounds and life experiences. With all of us being bi-/multilingual yet coming from varied cultures and backgrounds, we reflect a range of the diversity that comprises the communities we study, which predominantly speak more than one language. Indeed, as a group, we are diverse in our linguistic experiences. We come from various countries across Europe, Asia, and North America, with all of us speaking many languages (including minority, heritage and majority ones), acquired at different points in life, and some of us also raising multilingual children. In our research, we have all worked with diverse populations of both mono- and multilingual speakers. Our multilingual and multicultural experience, which is both shared across the group and unique to individual members, makes us particularly aware of the diversity and complexity of language experiences and how these differ between contexts, making them not fully comparable. While we acknowledge that we all bring our biases and assumptions to our work, including the present one, we have actively reflected on them in the course of writing this manuscript, which endeavors to deal with studying bi-/multilingualism more accurately and equitably.

Human (social) sciences relies heavily on the application of the standard scientific method for both quantitative and qualitative research (Whenwell, 2017). The construct of *methodological control* (METH-C) is a well-known, essential part of this method. A proper control can be defined as a procedure/practice/mechanism purposefully designed or otherwise

included to minimize the effects of confounding variables—those that can influence the (in)dependent variable(s) in experimental focus. Failing to have proper control reduces the meaningfulness of any found association. Without it, correlations are just as likely to be genuine or generalizable as they are (inadvertently) spurious or misleading. There is clear consensus: Experimental control is required in good (social) science. Empiricism in (applied) psycholinguistics is of course no exception. And yet, the implementation of METH-C in bilingualism research, typically based on monolingual comparative normativity, is (often) highly problematic and, in and of itself, can provide greater noise than it reduces.

Before unpacking this in greater detail below, let us consider a useful analogy in a context where standards for METH-C are potentially clearer, less personal (to those studying bilingualism), and its failure implies very high risk: The case of determining drug/vaccine efficacy in clinical trials. What is a control group in a clinical trial? A standard definition inevitably includes something like: *A group of participants who do not receive the drug/treatment/intervention under investigation, but instead receive a <u>standard of care</u> or a <u>placebo</u>* (Friedman et al., 2015). At least in (relatively) small-scale trials in medicine—the first step in a multi-phase process—matching of other potential intervening variables between the members of the intervention and non-intervention groups is essential to have a *bona fide* chance at revealing any meaningful effect from the drug/treatment itself. If an apparent effect emerges, subsequent phases include increasingly larger scale studies. These are meant to be representative of populations where variation in potential intervening variables is not (as tightly) controlled to tease out coverage in the real world: broadly representative effect sizes and, thus, generalizability. In such cases, the increased numbers, often in the tens of thousands, provide the necessary power to tease out variable interactions. For example, beyond the original question of *"Does a drug work at all?"*, one wants to know what its efficacy rates are and the limitations, if any, it has for certain populations or under certain conditions. In any case, crucial to the concept of METH-C is the promise that there is a potential for meaningful distinction between the compared entities: a positive effect in those receiving a targeted treatment and the lack of the hypothesized effect for those receiving nothing or a placebo. This works quite well in the world of pharmaceuticals. Yet can the same METH-C standards be applied in the social sciences, in particular in the world of bilingualism? Should they?

We will spend the entirety of this paper pondering and problematizing these questions. For now, suffice it to say, there is no reason, *a priori,* why they cannot apply. Yet, for various reasons in the domain of bilingualism, they have not. It seems to us, however, that not all researchers in bilingualism studies realize/understand that this is the case. In part, this is because determining how METH-C should be chosen or applied in bilingualism research— social sciences more generally—is not as straightforward as it (possibly) is in the world of STEM-focused research, requiring more diverse and numerous forms. In other words, there simply cannot be any METH-C default. And yet, the undeniably typical practice in bilingualism empiricism is one of a default monolingual comparative "*control*".

In this light, we wish to highlight an important issue that arises when considering the drug trial analogy. What are the parallel constructs for *the drug/treatment/intervention* and *standard of care* or *placebo* in bilingualism research? The first is relatively easy, we surmise most would say: *bilingualism* itself. Whether our questions are purely linguistic, educational, social, cognitive or at their interfaces, the goal of much bilingualism research is to understand what bilingualism as the *drug/treatment/intervention* implies/brings about. Yet, to apply the standard scientific METH-C we would have to be clear on what would constitute an analogous *standard of care* or *placebo* in our comparison groups. Doing so, however, is precisely where default monolingual normativity weakens. At least in the world of psycholinguistic/cognitive science approaches, *language* cannot—nor would anyone wish it to—be separated from bilingualism and any effect it might have in the focus of our experimental radar. As a result,

unlike the case of a clinical trial, there can be no *standard of care* that does not include language, nor is it clear what a *placebo* for it would be. Why? Because everyone has language(s) that include(s) multifarious and dynamic relationships with them. Language-specific and other language-related experience matters for <u>all</u>, be it for linguistic or cognitive outcomes, even and especially for individuals traditionally labeled as monolinguals (e.g., Dąbrowska, 2012; Street & Dąbrowska, 2010; Bice & Kroll, 2019). In fact, relatedly, discussions pertaining to the lack of clarity, inclusivity and utility of labels such as "monolingual" and "native speaker" and, indeed, questioning the "selectivity" of participants (monolinguals and bilinguals alike) that populate studies themselves have taken center stage in psycholinguistic research or criticisms thereof in recent years (e.g., Bayram, Kupisch, Pascual y Cabo & Rothman, 2019; Castro, Wodniecka & Timmer, 2022; Cheng et al., 2021; Flores, 2016; Ortega, 2020; López, Luque, & Piña-Watson, 2021; Rothman, 2008; Rothman & Treffers-Daller, 2014; Wiese et al., 2022). If there is no obvious *standard of care* or *placebo* in what constitutes the typical control group in bilingual studies—so-called monolinguals—then this typical control (often) fails in its most essential role because it is, in fact, controlling very little or nothing at all. Thus, the <u>default</u> nature of monolingualism as a METH-C has all the makings of a classical comparative fallacy.

Before delving any deeper, we wish to be clear from the outset that we are not suggesting monolinguals can never serve as suitable comparisons. Clearly, it would be just when the nature of a legitimate research question could <u>only</u> be addressed from it. For example, if one is interested in documenting the (potential) role that cross-linguistic influence has in the development of bilingual grammars in childhood, it could be reasonable to compare a child bilingual group to a monolingual one. With proper control of other potentially interceding variables that distinguish groups, distinctions between them could (and have) reveal(ed) the precise parameters that (bidirectional) cross-linguistic influence has for bilingual developmental sequencing and ultimate attainment. Let us further consider studies in clinical linguistics seeking to examine the efficacy of diagnostic measures of developmental language disorder applied to bilinguals. The by far typical case would be that such measurements were developed for, and normed with, monolinguals. Are they equally useful for monolingual and bilingual ones? Studies with these questions in mind have legitimate need for a monolingual comparison, by virtue of the questions they ask. Even so, in such cases, one needs to recognize that the monolingual group is a comparison which, while needed, does not fulfill the typical role of experimental control in the scientific method sense. Notwithstanding, the argumentation developed herein will lead to the conclusion that monolinguals cannot be taken for granted as an appropriate comparison, much less control. Rather, their use must be justified, case by case, in terms of its theoretical relevancy and ability to reduce confounding variables given the specific question(s) particular studies address. In other words, the utility of monolinguals as a control group, like any alternative, should be explicitly argued for. Alternatively, the message sent is that monolingualism is the "norm", if not gold standard, state of linguistic knowledge; it has some privileged status making it the ubiquitous target and/or benchmark; it does not naturally show significant levels of variation dependent on experiential factors, or that such variation is (theoretically) unimportant. Of course, none of this is true, defensible, or helpful to our shared goals as bilingualism researchers.

It is not simply neutral or innocuous to tolerate the superfluous presence of monolingual comparisons in studies whose research questions do not justify their inclusion. True for many reasons, the most pressing for the present context is that a monolingual aggregate's presence all but necessitates particular framings and interpretations that detract from the real value of bilingual data themselves. Perpetuating the default expectation of this comparison by allowing it—or never thinking to question it—regardless of its potential to meaningfully contribute to a theoretical query only it could speak to has elevated and sustained the perception of

monolingualism normalcy. Worse, in our view, being permissive in this way has contributed in no small part to a forestalling of progress towards understanding many aspects of bilingualism in their own right, in some cases delegitimizing its equality. Indeed, in much research that has included monolingual comparisons as a default, the inherent value of the bilingual data either remains inexhaustibly explored or is disingenuously contextualized. To offer one example from our own work, Bayram et al. (2019) included two monolingual comparison groups for their Turkish-German heritage language (HL) bilingual focus group. In that paper, we spent considerable space highlighting and discussing the bilinguals' use of passives (both Turkish and German) in terms of how they differed and/or were similar to German and Turkish "so-called monolingual controls". Ultimately, as we scrutinize in much greater detail in the final section below, where Bayram et al. (2019, 2020) are used as case studies, what comes of the HL bilingual to homeland speaker comparisons is not terribly interesting or informative. Alternatively, what emerges as really valuable is the studies' treatment of bilingual individual differences, isolating a key explanatory factor that contributed to bilingual intragroup variation. Ultimately, the monolingual comparison, descriptively interesting as some might find them, served to meet the default expectation of compulsory comparison, if not diverting some value away from the more important discoveries which could (should) have been done in the absence of monolingual groups.

Monolingual comparative normativity is especially unfortunate considering that bi-/multilingualism is actually the majority global reality (De Houwer, 2021). And so, given the context of our science and societies in 2022, it is worth bringing to the fore the problematizing of this practice. In the course of contextualizing our position further, we offer some theoretically sound alternatives for bilingualism METH-C that delegitimize the default status of monolingual comparative normativity while improving ecological validity, social justice prudence and scientific rigor simultaneously. It is our contention that if the exercise of having to justify a monolingual comparison in the way described above became the new norm a few things would ensue: (i) the realization that fewer studies actually need(ed) them, (ii) increased motivation to foster theoretically sound and socially responsible research that shifts the onus to where it has always belonged: understanding bilingualism and its effects in and of themselves, (iii) the emergence of exciting, new questions and trends in bilingualism research, and (iv) more research that embraces diversity and promotes moribund, endangered and historically un(der)represented languages and their communities that have fallen victim to monolingual comparative normativity.

Language science, broadly defined, encompasses many disciplines, from linguistics, education, psychology, and anthropology to neuroscience, sociology, and language pathology. While scientists bring the traditions of their own (sub)fields to bear, when our queries involve individuals where more than one language is at play, we converge in the world of bilingualism. As such, there are some concerns that are equally relevant for all bilingualism scholars. We believe that default monolingual normativity is a case in point that has deep-rooted causes and far-reaching implications for methodology and theory. While the present issue is not theory-free, it is ultimately theory-independent. And so, while we believe that the discussion until this point and what follows below apply more generally to all types of bilingualism, we limit the remainder of this article to heritage language bilingualism (HLB) for ease of exposition, on the one hand, and because the points we raise are particularly pertinent to it and made especially poignant by it. Of course, psycholinguistic studies of HLB can have a more linguistic or more cognitive focus, yet both suffer equally, in our view, from the same monolingual normativity bias. Hence, in suggesting alternatives, we will showcase linguistic and cognitive work involving HL bilinguals from our group for two reasons: (a) we have been as guilty as anyone in historically applying monolingual comparative normativity in our work, therefore, showing how doing so was unnecessary with self-awareness and criticism seems especially suitable and

(b) in recent years, we have attempted real strides in circumventing monolingual comparative normativity, which we hope can serve to highlight the usefulness of some alternatives we put forward.

Contextualizing the case of Heritage Language Bilingualism (HLB)

All HL bilingual individuals, henceforth *heritage speakers* (HSs)*,* are bilinguals, but not all bilinguals are HSs. Two obvious questions emerge from this statement: (i) what is a HL and (ii) what determines membership as a HS of one? While all languages have heritage links, in the present context HL refers to a specific sociolinguistic context where a given language is asymmetrically situated with respect to another (or other) language(s) in a given society where HSs are raised: *A language qualifies as a **HL** if it is a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society* (Rothman, 2009:156). Thus, all languages can be the dominant societal language in one context and a HL in another. Permutations of potential language pairings for HS contexts globally are simply massive. And yet, for various reasons, some of which also bear on issues of social justice and hegemonic imbalances of specific languages being disproportionately represented in language sciences, a majority of HS empirical studies focus on a very small set of languages and pairings: English is overly represented as the dominant societal language, and relatively few HLs make up the majority of the literature (Scontras & Putnam, 2020). Obviously, the field needs increased diversity in the represented HLs and more variety in specific-language pairings and contexts in which they are found. This is true for several reasons simultaneously. It would aid in better leveraging and elevating HLB as a natural laboratory for testing the utility and unique contributions of HLB data for hypotheses, constructs and descriptions in formal linguistic, language acquisition and processing theories (e.g., see Lohndal et al., 2019). It would also increase ecological validity and potential for generalizability: At present not having diverse representation in HLB studies means that we have access to only a fraction of the contexts and potential data one needs to claim truths and generalizations about HLB *par excellence*. And while HSs are, by definition, minorities in the majority language context in which they are studied, expanding linguistic diversity will surely mean that more minoritized peoples, lesser commonly spoken languages, and some languages historically perceived as less prestigious will come to have more meaningful representation.

(Psycho)linguistic studies of language acquisition and processing in HLB have progressively grown in number and sophistication over the past three decades. With the caveats of holes in coverage and representation discussed above, what does the composite of this research tell us? A primary focus of HLB research relates to the differences that HSs show when compared, most typically, to monolingual counterparts or, much less often, to dominant native-speakers of the HL who are also bilinguals (e.g., HSs of a different language or second language speakers of another language) but grew up where the language in focus is the main societal one (e.g., Montrul, 2008; Polinsky, 2018). Implicitly or explicitly, these comparison groups, often referred to in writing as the "*control* group", are taken as the HS comparative baseline (e.g., Polinsky 2018; Polinsky & Scontras, 2020). This is so, despite the fact that the conditions under which the language of comparison—the HL—was acquired, is distributed, and used are likely to be significantly distinct across potentially innumerable axioms. Going back to the example of the clinical trials, it seems as if heritage speaker-hood is taken as the *drug* whose effect is being examined as if it comprised a monolithic entity tested against the outcomes of a *standard of care group* who gets no HLB intervention. What the groups have in common is that both are native speaking naturalistic childhood learners of what is taken to be the same (heritage) language (Rothman & Treffers-Daller, 2014). Yet, crucially to our main point, so much else varies—in our analogy, the chemical composition of the *drug* itself—

between the groups and across the individuals that comprise them: quantities and qualities of input, opportunities for converting input into intake, the (lack of) opportunities for formal training in the HL, the social milieu and distribution of language use, to say nothing of the fact that HSs are bilingual whereas many so-called baselines are not. And so, if HLB is the *drug* being tested, it is troubling that individuals in the treatment group can receive vastly different *drugs* in actuality only to be compared against a *standard of care* group who *supposedly* received no *drug* at all, as if differences in the chemical composition of the *drug* do not matter in the aggregate. As problematic as it would be to gauge efficacy of "drugs for back pain" by knowingly combining participants receiving Aspirin, Acetaminophen, Ibuprofen and Codeine as a single treatment group, compared to a *standard of care group* receiving none, it is equally unfitting to effectively do this in bilingualism research (without having proper procedures in place to leverage this, e.g., seeking to run *drug* type as a regressor variable with questions that justify this).

Indeed, individual HS differences to so-called baselines present across a considerable spectrum, as can monolinguals to one another. While a study might claim to show that HSs have a "reduced" grammar for a particular domain on average compared to the assumed baseline, e.g., a two-way gender system in HL Russian in the US compared to the three-way gender system of monolinguals in Russia (Polinsky, 2008)—looking at the individual HSs in these same studies often reveals some show no such innovations: They are experimentally indistinguishable from the assumed baseline. Thus, HSs themselves present on a spectrum of distinct linguistic outcomes from one another. Research that probes into what variables explain HS-to-HS differences shows systematicity in accounting for them, often related to engagement opportunities with the HL such as formal literacy exposure or the compositionality of the home environment (Bayram et al., 2019; Lloyd-Smith, Bayram & Iverson, 2020). And yet, the typical description of HSs is that, insofar as they differ from the default baseline, they are "unbalanced" bilinguals whose grammars and processing differences in the HL reflect "reduced", "arrested", "incomplete" and/or "simplified" systems.

We take it for granted that any uses of these above modifiers/labels are intended—as is the case when we have used them in the past—as nonevaluative descriptors, that is, without any implicit prejudice implied. However, some researchers maintain that these labels have irreplaceable value presumably because they take HS grammatical differences to, in fact, reflect incompleteness even while acknowledging their grammars are rule-governed, highly systematic and universally compliant (Domínguez, Hicks & Slabakova, 2019). In our view, such a position and the very existence of these labels could only exist in a context of monolingual comparative normativity. Afterall, the very terms "reduced", "incomplete" and/or "simplified" are relative ones, for example, the very semantics of the word "incomplete" entails a measure of potential completeness that is left unattained. Putting aside for the moment that monolingual individuals also show (constrained) variation, were it the case that monolinguals are—always—an appropriate comparative control and indeed the appropriate baseline for HS knowledge and performance, then we suppose it would be less egregious to label deviations from that benchmark as incomplete because the target of completion is set by the baseline itself. But if the baseline is not actually the baseline to which (many, most) HSs actually have exposure, then labeling differences as incomplete is misleading, if not theoretically vacuous. Interested as one might be in the comparative differences that are viewed as evidence of incompleteness, such differences alone tell one nothing about why and how they came to be, much less highlight any immediate theoretical significance of such descriptions at a higher level (Bayram, Kupisch, et al., 2019).

What, then, explains the spectrum of outcome differences in HSs from often ill-conceived baselines as well as to one another? Unlike other native speakers of a given language growing up in a native-dominant society, the continuum of relevant variation that HS

individuals are likely to experience is larger. Furthermore, there is (virtually) no guaranteed access to a particular standard and formal training in it—as is the case for monolingual groups in typical studies—to obscure idiolectal and dialectal distinctions flying under the radar in psycholinguistic testing. HSs demonstrate that native outcomes can be much more varied than what psycholinguistics has typically found (and has perhaps been comfortable with showing). This, of course, cannot be separated from other co-occurring issues related to who typically populates psycholinguistic studies in terms of demographic, cultural, ethnic, educational and SES backgrounds. To take one variable, as already alluded to, HSs often lack any formal training in a standard variety of the HL, yet so-called monolingual control participants who populate psycholinguistic studies almost always have significant education in it. This alone introduces noise and renders the comparison uncontrolled. When one accounts for this, for example, by testing HSs who are formally educated in the HL—rare as this actually is done—to an equal degree (e.g., French HSs growing up in Germany, are German dominant but have attended French medium schools), then differences between (functional) monolinguals greatly reduce, if not disappear entirely (Kupisch & Rothman, 2018). All of this leads to the conclusion that comparing typical HSs to monolinguals has limited value, not least since often used comparative constructs such as *proficiency* typically underlyingly assume a (particular) monolingual baseline-as-benchmark whose convergence potential entails extraneous conditions such as access to literacy and/or particular registers.

Since the majority of empirical work in HLB has not been questioned and thus followed default *monolingual comparative normativity*, the question now turns to what should replace it. We submit that comparing HSs to other types of bilinguals and understanding HS-to-HS individual variation are more equitable and theoretically interesting avenues to pursue, not least because one can more meaningfully probe into the variables that conspire to result in documented variation itself. Crucially, it removes the comparative fallacy of using monolingual baselines. It embraces the ubiquity and normalcy of bilingualism while recognizing the determinism of linguistic, cognitive and societal variables that are either unique to bilingualism or distribute differently within and across bilinguals. Because inter-participant variation in HS aggregate compositions on key variables—e.g., larger and smaller social networks, degree of literacy and formal education in the HL, patterns of use at home, at work, in social contexts—can be run as predictive variables in regression models, such an approach will help to legitimize all language contexts and all individuals as being equally worthy of investigation. Unencumbered by the default need for a monolingual control, such an approach will open up avenues for promoting a larger array of minoritized individuals with distinct profiles (not mainly ones available in psychology pools at our universities). For example, languages that do not have monolingual communities from which a so-called baseline control can be formed can be more easily investigated in relevant literatures. All of this culminates in seeing the bilingual forest for the trees: There is much value in understanding (HL) bilingualism independently.

There are many ways to deal with default monolingual comparative normativity. Here we wish to highlight three alternatives that our group has focused on in recent years: (i) studies where there is no (reason for an) aggregated comparative analysis control/comparison group, (ii) studies that use other bilinguals as a comparative group to HSs where the two languages in the pairing are held constant, for example, studies where HSs are compared to potential L1 attritors of their HL (the community who provided the HSs' primary linguistic input, e.g. Pascual y Cabo, 2020) and (iii) studies that compare various language pairings where the HL is held constant (ideally HSs are matched on other key demographic and experience variables), but the majority languages differ to test for constraints on cross-linguistic influence effects and how they interact with/are conditioned by (differences in) HL context (e.g., van Osch, 2019). Before delving any deeper, we would be remiss to not acknowledge that (i) to (iii), but

especially (ii) and (iii) could be viewed as potential imposers of practical constraints that narrow the populations and questions that can be asked, i.e., as types of restrictive gatekeepers. Given that excessive gatekeeping is one of the key contributors to default monolingual comparative normativity in the first place, if so this would be less than ideal to say the least. As such, it is important to understand (i) to (iii) in their proper contexts. Firstly, we do not believe they pose insurmountable obstacles, although surely they are not all equally plausible in every context. But more importantly, they are decisively not intended to be the solutions to a problem *per se.* Rather, they are merely three examples of ways our group has attempted to maintain rigorous empirical standards while rejecting the default monolingual comparative *status quo*. Indeed, the list of alternatives herein is incomplete, yet we have found these mere examples useful. Thus, we encourage them when appropriate and useful as much as we welcome further suggestions to be added to a more comprehensive list. Space limitations for this paper being what they are and not least because we find (i) the most insightful, scalable and inclusive, we walk the reader through (i) in greater detail below.

Using Bilingual Experiences as Regressors
The first alternative suggested above presents a context in which no comparison group is included (or needed) in the design; a shift away from aggregate comparisons towards unpacking individual differences. This can be done in several ways. For example, one could collapse traditionally separated groups (i.e., monolinguals and/or various types of bilinguals), to a single larger one where the type of (mono/bi/multi)-*lingualism* is regressed as a variable of interest and ideally interacted with other predictive variables for individual differences. At first glance, this might appear to be more of the same: an *a priori* assumption about distinctions between monolinguals and bilinguals. However, it is not at all. The very impetus of doing it this way is because one is open to potential overlap across the traditionally separated group members and is committed to unpacking the relevance of this: some monolinguals will perform like bilinguals and vice versa even if they seem to be distinguishable, by averaging, into respective aggregates. Underlying this approach is the hypothesis that individual-level quantities, qualities, usage patterns and contextual opportunities with language(s)-related experience ultimately predict individual outcomes regardless of -*lingualism* type. Applying this type of approach and in line with a general trend in the neurocognition of bilingualism literature looking to understand bilingualism as a reflection of the spectrum of its experiences (e.g., DeLuca et al., 2019), a recent study by Pereira Soares and colleagues (2022) examined the neural signatures of cognitive control by collapsing adult sequential bilinguals (highly proficient second language learners) of English in Germany with HSs of Italian in Germany. Using electroencephalography (EEG), they examined the oscillatory dynamics—rhythmic patterns of neural activity— of inhibitory control measured in a Flanker task. At the collapsed group level, age of acquisition of the non-societal language and usage of it specifically at home (the extent to which English and/or Italian is used at home) positively predicted low beta recruitment on the task. What remained unclear, however, was if bilingual type was an independent driving factor for the combined aggregate results. In other words, was a particular subgroup of bilinguals, e.g., HSs speaking more Italian at home than L2 learners would be speaking English, skewing overall applicability? In a second phase analysis, the groups were separated by bilingual type. Qualitatively, both groups exhibited similar neural signatures. However, examining individual differences revealed that the type of bilingualism mattered in how these signatures were modulated by experiential factors. In other words, individual differences predicted oscillatory dynamics of inhibitory control across all bilinguals, but differentially so. For example, the duration of exposure to the non-societal language (as proxied by the chronological age at the time of testing) in the group of proficient second language learners predicted attentional engagement during the cognitive control task (as reflected in

alpha power), while the same variable in the group of HSs predicted conflict resolution demands on task (as reflected in theta power). This suggests that above and beyond degree of engagement with bilingual experience, there is something additional to age of onset to bilingualism that interacts to make the confines of bilingual-related individual differences in neurocognitive effects distinct. And so, while Pereira Soares et al. provide interesting data showing that bilingual effects on neurocognition obtain on a predictable continuum in general, the data also highlight how the brain adapts differentially even to similar degrees of bilingual language experience depending on when, within the age-related developmental continuum, a person first becomes bilingual. Thus, while interesting trends can be appreciated via collapsing bilinguals of various types, we wish to also highlight two additional things: (a) applying such an approach does not mean—assuming statistical power is present—that one has to sacrifice subsequent group comparative analyses when warranted and (b) there can be added value insofar as collapsing permits one to see where individual bilinguals fall along articulated continua of overlapping variables between traditionally separated groups—some HSs and L2 learners were indeed indistinguishable—while appreciating still that HSs should be treated distinctly in bilingual neurocognitive research.

Another approach that sidesteps the perceived need for a (monolingual) comparative control would be to have (large) cohorts of bilinguals of one specific type—say, only HSs—in an analysis where data related to individual bilingual experiences—ones quantifying variables hypothesized to matter for a given effect of focus—are regressed to probe for individual differences. Such an approach prioritizes understanding bilingualism itself and promises answers as to how and why individual variation obtains the way it does. Let us consider a few relevant studies from our lab in light of the above. Bayram et al. (2019) and Lloyd-Smith et al. (2020) examined Turkish HSs who grew up in Germany and were thus dominant speakers of German. Bayram et al. (2019) examined the morphosyntax of passive voice under an experimental paradigm of constrained elicited production. Participants were given 14 sets of four pictures in which the final two were accompanied by patient-focused questions (i.e., what is happening to the little fish?) under the hypothesis that such a question frame would augment the naturalness for a passive response. Four experimental data sets were provided: two so-called control groups (a monolingual Turkish group tested in Turkey, a monolingual German group tested in Germany) and data from the same HSs for both Turkish and German. Passives were chosen for several reasons, not least because of how the structure presents in the two languages. In brief, German essentially works like English but Turkish, an agglutinative language, is quite distinct. Turkish has dedicated voice morphology. Passives are formed by adding the morpheme -il, or an appropriate allomorph, to the verbal stem. The distribution of overt case marking in passive constructions can be affected: (specific) Turkish direct objects are overtly accusative-marked while matrix subjects are nominative-marked with a null exponent. These Turkish-specific facts presented a (relatively) unique opportunity to see if Turkish HSs have the morphosyntax of passives at all. To show this, they would need to provide a morpheme that was absent within the experimental procedure. The authors compared the four data sets in the traditional way as a first pass: The Turkish HS aggregate performed statistically indistinguishably from the so-called German controls in German and significantly differently from the Turkish monolinguals in Turkish. Yet literally every single one of the HSs produced at least one passive in Turkish and in all cases that a passive was given, the entirety of the structure (case alternations) was grammatical. And so, what can be concluded?

The data showed that these HSs of Turkish have the morphosyntax of passives but use it much less frequently in this experiment as compared to the Turkish monolinguals. Interestingly, suppliance of passives by the HSs in German and Turkish was not distinct from which one might conclude that while they have the underlying structure in each language, their dominant language, German, exercises some influence on how passives are deployed in

Turkish. But was this true of all Turkish HSs or is this simply an impression by looking at the aggregated data? Looking at the range (suppliance = 1-13) of individual performances in Turkish, it was clear that the average of 7 passives provided was not representative of all individuals. Bayram and colleagues did a further analysis only on the HS data both for the German and Turkish productions. Applying logistic regression analysis, they regressed age at the time of testing, parental background (immigration status of both parents) and exposure to formal Turkish literacy (none, self-taught, less than 3 years in Turkish supplementary school or 5 years or more). As it turned out, only one variable mattered: Turkish literacy scores provided a wonderful fit of the data. Crucially, there was no overall effect, meaning, it only predicted individual differences in the Turkish data (i.e., not the German of the same participants).

So, what did the first pass analysis tell us? Turkish HSs are different than monolinguals. What independent value does knowing that have, especially since hundreds of studies have shown similar differences? We would say it tells us little because it is ultimately descriptive. What explains inter-HS variation is what really mattered because it was highly systematic and predictive: It addressed straightforwardly the why and how. As it turns out then, we did not really need the so-called controls. Not only were they not really controlling much anyway, the comparison itself did not do any justice to the data as a whole. And to examine the inter-HS differences, their presence was superfluous. To the potential (yet fallacious) point that the monolinguals would have provided added value in the sense of quality control in the experiment—that they produced passives means the experiment worked—again they were superfluous. The HSs too produced passives thus we know the experiment worked from their data alone.

In a follow up study, Lloyd-Smith et al. (2020) looked at narrative elicitation of the exact same participants collected at the same time performing the *Frog Where are you?* story *(*Mayer, 1969*)*, this time with no so-called control groups. They calculated two holistic measures from the productions: type-token frequency (TTR) as a measure of lexical richness and a measure of clausal morphosyntactic complexity (cMSC). Like the previous study, there was significant variation across the HSs. Running the same variables as regressors, there was again systematicity in coverage of individual differences for both measures in Turkish. However, the variable this time was different: In both cases it was parental background. For both TTR and cMSC it did not matter if one was 8 or 12 at the time of testing, nor whether or not they had any formal training in Turkish but indeed whether or not both, one or none of their parents were immigrants themselves (as opposed to 2nd generation HSs) of Turkish. Bringing these two studies together highlights the complexity of what underlies individual differences in HSs. There is not one variable that is explanatory for every domain of difference in HSs and stopping at or being satisfied with descriptively documenting differences misses several crucial marks. In fact, doing so would make it impossible to address this complexity and reveal this underlying systematicity of HSs in the first place.

Conclusion: Wrapping Things Up
While everyone agrees that METH-C is required in empirical bilingualism work, our main goal here was to take stock regarding how this has been traditionally applied and provide a critical discussion that asks the field to (re)-consider: (i) the form(s) controls should actually take, (ii) when some are more and less appropriate and/or (iii) all applicable confounding considerations for determining/applying them in line with their ability/suitability for specific questions. In doing so, we honed in on merely one METH-C construct we consider especially perplexing and troublesome in the study of bilingualism: the seemingly default expectation of monolinguals as "the baseline". And while we placed a specific spotlight on HLB, as mentioned in the outset, the take-home messages and the crux of the offered argumentation

apply to all instances of bilingualism. As discussed, there is no question that monolingual to bilingual comparisons have been and can continue to be fruitful and theoretically relevant. Yet, in our view, the juxtaposing of monolingualism against bilingualism has been too far-reaching. Doing so has contributed to the sweeping under the rug of (inherent) confounds that unnecessarily compromise a general understanding of the full extent to which bilingual data make independent contributions to psycholinguistics and cognitive (neuro)science. Since it is likely the case that default monolingual comparative normativity has limited the set of questions we currently even know to ask in the field, we are hopeful that problematizing its *status quo* nature and promoting the field's eventual departure from it will set the stage for new, more inclusive questions in the near future. As it would be disingenuous to deny the relationship between empirical monolingual comparative normativity and the perception of monolingualism-as-normalcy in Western cultures, we also expect that dealing with it would not only improve empiricism in bilingualism research, but also act as a counterbalance to the practices that have resulted in the lack of diversity and representation in psycholinguistic studies to date.

References

Bayram, F., Rothman, J., Iverson, M., Kupisch, T., Miller, D., Puig-Mayenco, E., & Westergaard, M. (2019). Differences in use without deficiencies in competence: passives in the Turkish and German of Turkish heritage speakers in Germany. *International Journal of Bilingual Education and Bilingualism*, 22 (8), 919-939

Bayram, F., Kupisch, T., y Cabo, D. P., & Rothman, J. (2019). Terminology matters on theoretical grounds too!: Coherent grammars cannot be incomplete. *Studies in Second Language Acquisition*, *41*(2), 257-264.

Bice, K., & Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language*, *196*, 104644.

Castro, S., Wodniecka, Z., & Timmer, K. (2022). Am I truly monolingual? Exploring foreign language experiences in monolinguals. *PloS one*, *17*(3), e0265563.

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.715843

Dąbrowska, E. (2012). Explaining individual differences in linguistic proficiency. *Linguistic Approaches to Bilingualism*, *2*(3), 324-335.

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, *116*(15), 7565-7574.

De Houwer, A. (2021). *Bilingual development in childhood*. Cambridge University Press.

Domínguez, L., Hicks, G., & Slabakova, R. (2019). Terminology choice in generative acquisition research: The case of "incomplete acquisition" in heritage language grammars. *Studies in Second Language Acquisition*, *41*(2), 241-255.

Flores, N. (2016). A tale of two visions: Hegemonic whiteness and bilingual education. *Educational Policy*, *30*(1), 13-38.

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., & Granger, C. B. (2015). *Fundamentals of clinical trials*. Springer.

Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, *22*(5), 564-582.

Lloyd-Smith, A., Bayram, F., & Iverson, M. (2020). The effects of heritage language experience on lexical and morphosyntactic outcomes. In Bayram, F. (ed.) *Studies in Turkish as a Heritage Language*, *60*, 63-86. John Benjamins Publishing.

Lohndal, T., Rothman, J., Kupisch, T., & Westergaard, M. (2019). Heritage language acquisition: What it reveals and why it is important for formal linguistic theories. *Language and Linguistics Compass*, *13*(12), e12357.

López, B. G., Luque, A., & Piña-Watson, B. (2021). Context, intersectionality, and resilience: Moving toward a more holistic study of bilingualism in cognitive science. *Cultural Diversity and Ethnic Minority Psychology*. doi: http://dx.doi.org.mime.uit.no/10.1037/cdp0000472

Mayer, M. (1969). *Frog, where are you?*. New York: Dial Press

Montrul, S. (2008). *Incomplete acquisition in bilingualism. Re-examining the age factor*. John Benjamins Publishing.

Ortega, L. (2020). The study of heritage language development from a bilingualism and social justice perspective. *Language Learning*, *70*, 15-53.

van Osch, B. (2019). Vulnerability and cross-linguistic influence in heritage Spanish: comparing different majority languages. *Heritage Language Journal, 16*(3), 340-366.

Pascual y Cabo, D. (2020). Examining the role of cross-generational attrition in the development of Spanish as a heritage language: Evidence from gustar-like verbs. *Linguistic Approaches to Bilingualism*, *10*(1), 86-108.

Pereira Soares S.M, Prystauka Y., DeLuca V., & Rothman J. (2022). Type of bilingualism conditions individual differences in the oscillatory dynamics of inhibitory control. *Frontiers in Human Neuroscience*. 16:910910. doi: 10.3389/fnhum.2022.910910.

Polinsky, M. (2018). *Heritage languages and their speakers* (Vol. 159). Cambridge University Press.

Polinsky, M., & Scontras, G. (2020). A roadmap for heritage language research. *Bilingualism: Language and Cognition*, *23*(1), 50-55.

Rothman, J. (2008). Linguistic epistemology and the notion of monolingualism. *Sociolinguistic Studies, 2*(3), 441-458.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International journal of bilingualism*, *13*(2), 155-163.

Rothman, J., & Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too!. *Applied linguistics*, *35*(1), 93-98.

Scontras, G., & Putnam, M. T. (2020). Lesser-studied heritage languages: An appeal to the dyad. *Heritage Language Journal*, *17*(2), 152-155.

Street, J. A., & Dąbrowska, E. (2010). More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers?. *Lingua*, *120*(8), 2080-2094.

Whewell, W. (2017). Comte and positivism. *Epistemology & Philosophy of Science*, *54*(4), 209-224.

Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., ... & Zuban, Y. (2022). Heritage speakers as part of the native language continuum. *Frontiers in Psychology*, 12:717973