Routledge
Taylor & Francis Group

Aging, Neuropsychology, and Cognition

A Journal on Normal and Dysfunctional Development

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/nanc20

Effects of increasing fitness through exercise training on language comprehension in monolingual and bilingual older adults: a randomized controlled trial

Eunice G. Fernandes, Sindre H. Fosstveit, Jack Feron, Foyzul Rahman, Samuel J. E. Lucas, Hilde Lohne-Seiler, Sveinung Berntsen, Allison Wetterlin, Katrien Segaert & Linda Wheeldon

To cite this article: Eunice G. Fernandes, Sindre H. Fosstveit, Jack Feron, Foyzul Rahman, Samuel J. E. Lucas, Hilde Lohne-Seiler, Sveinung Berntsen, Allison Wetterlin, Katrien Segaert & Linda Wheeldon (18 Dec 2024): Effects of increasing fitness through exercise training on language comprehension in monolingual and bilingual older adults: a randomized controlled trial, Aging, Neuropsychology, and Cognition, DOI: 10.1080/13825585.2024.2435914

To link to this article: https://doi.org/10.1080/13825585.2024.2435914

AGING, NEUROPSYCHOLOGY, AND COGNITION
https://doi.org/10.1080/13825585.2024.2435914

Effects of increasing fitness through exercise training on language comprehension in monolingual and bilingual older adults: a randomized controlled trial

Eunice G. Fernandes[a], Sindre H. Fosstveit[b], Jack Feron[c], Foyzul Rahman[d], Samuel J. E. Lucas[c], Hilde Lohne-Seiler[b], Sveinung Berntsen[b], Allison Wetterlin[a], Katrien Segaert[d] and Linda Wheeldon[a]

[a]Department of Foreign Languages and Translation, University of Agder, Kristiansand, Norway; [b]Department of Sport Science and Physical Education, University of Agder, Kristiansand, Norway; [c]School of Sport, Exercise, and Rehabilitation Sciences, University of Birmingham, Birmingham, UK; [d]School of Psychology, University of Birmingham, Birmingham, UK

ABSTRACT

Exercise training has been proposed to counteract age-related cognitive decline through improvements in cardiorespiratory fitness (CRF hypothesis). Research has focused on cognitive domains like attention and processing speed, and one cross-sectional study reported a positive relationship between CRF and language production in older adults. In a randomized controlled trial, we investigated whether these benefits could extend to language comprehension in healthy older adults, and whether bilinguals, for whom language processing is more costly, would exhibit greater benefits than monolinguals. Eighty older English monolinguals and 80 older Norwegian-English bilinguals were randomized into either a 6-month exercise training group or into a passive control group. We assessed CRF ($VO2_{peak}$) and language comprehension (reaction times to spoken word monitoring) in first (L1, all participants) and second language (L2, bilinguals only), before and after the intervention. We found that monolinguals in the exercise group (compared to the control group) were faster in comprehension following the intervention. Moreover, this effect was mediated by exercise-induced increases in $VO2_{peak}$, supporting the CRF hypothesis. This extends previous cross-sectional research and establishes a causal link between exercise training and speeded comprehension in older monolinguals. However, despite inducing increased $VO2_{peak}$, exercise training did not affect bilingual (L1 or L2) comprehension, and bilinguals in both groups were slower after the intervention period. Exploratory analyses suggested that this slowing may be driven by participants with low L2 proficiency, but further research is needed to examine whether bilingual language processing is in fact unaffected by exercise training and its consequent improvements in CRF.

ARTICLE HISTORY
Received 14 May 2024
Accepted 22 November 2024

KEYWORDS
Exercise training; cardiorespiratory fitness; language comprehension; bilingualism; ageing

CONTACT Eunice G. Fernandes e.fernandes1977@gmail.com Department of Psychology, University of Minho Escola de Psicologia - Edifício 14, Universidade Do Minho, Campus de Gualta, Braga 4710-057

Present affiliation of Eunice G. Fernandes is Department of Psychology, University of Minho, Braga, Portugal

Supplemental data for this article can be accessed online at https://doi.org/10.1080/13825585.2024.2435914

Introduction

Physical exercise, cognition, and language processing in ageing

Regular exercise is thought to offer protection against the adverse effects of aging for cognitive functions such as attention and processing speed, memory, and executive function (see e.g., Smith et al., 2010; Northey et al., 2018, for meta-analyses). According to the cardiorespiratory fitness hypothesis (Voss & Jain, 2022), such benefits of exercise are mediated by physiological adaptations associated with improvements in cardiorespiratory fitness (CRF). CRF (also termed aerobic fitness) is the capacity of the cardiovascular and respiratory systems to supply oxygen-rich blood to the working skeletal muscles and the capacity of the muscles to use oxygen to produce energy for movement and is most commonly indexed by maximal oxygen uptake ($VO2_{max}$ or $VO2_{peak}$; e.g., Caspersen et al., 1985).

Segaert et al. (2018) hypothesized a relationship between CRF and language processing, as the brain regions that show structural integrity differences associated with CRF (Haeger et al., 2019; Kramer & Erickson, 2007; Voss et al., 2013) overlap with the regions showing functional activation during language processing, in comprehension and production, namely, the frontal and temporal lobes (e.g., Menenti et al., 2012; Tyler et al., 2010). Indeed, Segaert et al. (2018) demonstrated an association between higher aerobic fitness and decreased tip-of-the-tongue (ToT) incidence in monolingual older adults and this, to our knowledge, remains the only study to have examined the relationship between CRF and language performance.

Word finding difficulties arising as ToT states are one of the most frequently reported language failures associated with aging (Maylor, 1990; Ossher et al., 2013). These and other age-related difficulties in speech production, such as longer pauses and slower production rates, are well documented (e.g., Duchin & Mysak, 1987; see Peelle, 2018, for a review). However, aging is also associated with some decline in language comprehension, an important ability for older adults to maintain social relations and independence. For example, older adults are slower and less accurate than younger adults in comprehension of written and spoken sentences. Age-related differences in comprehension are most often observed for complex sentences (e.g., Caplan et al., 2011) and "offline" measures (i.e., post-interpretive processes that occur after the meaning has been determined, such as grammaticality judgments, which introduce task-related cognitive demands; e.g., Poulisse et al., 2019). However, they have also been observed for the "online" processing of sentences without syntactic complexities (cf. Fernandes et al., 2024; see; Abrams & Farrell, 2011; Peelle, 2018, for reviews). Therefore, language comprehension seems to be negatively affected by healthy aging in ways that cannot be fully explained by the complexity of the linguistic materials or decline in domain-general cognitive systems (we will return to this issue in our Discussion).

The primary objective of our present study was to investigate whether exercise training can ameliorate language comprehension decline in healthy older adults. Crucially, cross-sectional designs such as Segaert et al. (2018) cannot establish a causal link between CRF and linguistic performance. In the present randomized controlled trial (RCT), participants were randomly assigned to either a passive control or a six-month exercise intervention group, enabling causal inferences between exercise training and linguistic performance.

Despite a general consensus and evidence for a positive relation between aerobic fitness and cognitive function (e.g., Colcombe et al., 2004), the physiological adaptations that slow cognitive aging are not yet completely understood (Voss & Jain, 2022; we return to this issue in the Discussion). Therefore, we also conducted mediation analyses in order to quantify and test the significance of one metric of CRF, peak oxygen consumption ($VO2_{peak}$), as a mediator of the effects of exercise training on linguistic performance.

Bilingualism and language comprehension

Our second main objective was to examine whether exercise training had different impacts on monolingual and bilingual language comprehension. Bilingualism incurs language processing costs that often surface as slower language processing compared to monolinguals, even when bilinguals use their first language (L1). For example, bilinguals are slower to name pictures, and look longer at written words while reading, compared to monolinguals (Gollan et al., 2011). In addition to this "bilingual disadvantage," there have been reports of "language dominance effects" whereby second language (L2) processing is more effortful than processing of the first or dominant language (e.g., Duyck et al., 2008; Gollan et al., 2011). It is possible, therefore, that any protective effects of exercise for language function could be even more important for bilinguals than monolinguals and, potentially, more important for L2 than L1 comprehension. Therefore, in healthy older adults, we tested L1 comprehension in mono- and bilinguals, and also L2 comprehension in bilinguals.[1]

Whereas the bilingualism and language dominance effects are well established, it is less clear what the causes of these disadvantages are. One aspect research has focused on is the specific domains of language processing where disadvantages are observed. Gollan et al. (2011, Experiment 2), for example, had young English monolinguals and (English dominant) Spanish-English bilinguals read English sentences with target words (e.g., "apple") embedded in low- or high-constraint sentences, that is, lacking or containing a semantic context that makes the target word predictable (e.g., "The artist painted the bowl and the apple" or "Snow White ate a poisoned apple"). Eye movements were recorded and the results showed that gaze durations to target words were shorter on high- compared to low-constraint contexts, and in monolinguals compared to bilinguals, but these effects did not interact, suggesting that bilingualism did not affect the use of semantic information provided by the sentence constraint. However, bilinguals can be at a disadvantage in the use of syntax to build the structure of sentences, as shown by Fernandes et al. (2024). They found that bilinguals were slower than monolinguals to detect words (e.g., "spatula") embedded in low-constraint sentences (i.e., with a normal structure; e.g., "I tried to quickly find the spatula to flip the pancake without breaking it") compared to random word lists (i.e., non-structured word lists; e.g., "Tried I find to quickly the spatula without pancake it to flip the breaking"), suggesting more difficulty in using syntactic information to guide comprehension.

Concerning language dominance, research suggests that syntactic processing may be more effortful in L2 than L1 processing. Several studies examined syntactic processing using complex structures such as garden path sentences like "Put the frog on the napkin into the box," where comprehenders initially parse "the napkin" as the goal (i.e., endpoint) of the action of putting, but then have to revise this incorrect interpretation, in face of the correct goal of "put," "in the box." Pozzan and Trueswell (2016) asked participants to hear alike sentences and act on them by moving corresponding objects (e.g., "frog," "napkin," "box") and found that L2 speakers of English made more errors than monolinguals, reflecting costlier recovery from the incorrect initial analysis. On the other hand, it is less clear whether semantic processing is prone to dominance effects. In the aforementioned study, Gollan et al. (2011) also compared the (English dominant) Spanish-English bilinguals to Dutch-English bilinguals, for whom the language tested (English) was L2. Here too, no interaction was found between the group of speakers and the contextual constraint manipulation (low- vs. high-constraint), suggesting that bilinguals' comprehension of L1 and L2 was not differently affected by the availability of constraining semantic information. However, there is contrasting evidence from studies measuring brain responses. Martin et al. (2013) measured the amplitudes of the N400 (a negative brain signal observed for semantically incorrect sentence endings, as compared to semantically correct ones) when L1 and L2 comprehenders read sentences with expected or unexpected endings like "Since it is raining, it is better to go out with an umbrella" or "She has a nice voice and always wanted to be an artist." They observed that the N400 effect was larger in L1 than L2 comprehension, suggesting that L2 comprehension relies less on semantic information than L1 comprehension. Therefore, how semantic processing differs in L1 and L2 remains an open question, and studies using different methods do not converge on their findings (we return to this issue in the Discussion).

In order to address these questions, we assessed language comprehension in a task tapping into effects of syntactic and semantic processing, i.e., of structure and meaning. We used a word monitoring task (Marslen-Wilson & Tyler, 1975, 1980; Tyler et al., 2010) where participants listened to spoken sentences and made a speeded button press when they heard a pre-specified target word. The target (e.g., "spatula") could be embedded in lists of words in random order (RWO; e.g., "Tried I find to quickly the spatula without pancake it to flip the breaking"), low-constraint (e.g., "I tried to quickly find the spatula to flip the pancake without breaking it"), or high-constraint sentences (e.g., "I flipped the pancake with the spatula without breaking it"). This task reflects the online construction of linguistic representations on which comprehenders build expectations for upcoming words. In low-constraint sentences, but not in RWO, comprehenders can build coherent syntactic representations. In high-constraint sentences, the semantic information constraints interpretation more than it does in low-constraint sentences. Therefore, differences in reaction times (RTs) to detect target words between RWO and low-constraint conditions index the use of syntactic information, whereas differences between low- and high-constraint sentences index the use of semantic information.

Current study

Monolingual and bilingual older adult speakers were randomly assigned to either a 6-month high-intensity interval training (HIIT) exercise group (exercisers) or to a passive control group (controls). HITT involves short bouts of intense exercise (i.e., >80% of maximal heart rate) interspersed with periods of recovery and has been shown to effectively improve CRF in older adults (Bouaziz et al., 2020). Before and after the intervention, CRF (i.e., $VO2_{peak}$) and language comprehension in L1 (all participants) and L2 (bilinguals only) were assessed. We hypothesized that exercise training would improve language comprehension in older adults. Specifically, exercisers would be faster (i.e., shorter latencies) to detect words in spoken sentences following the intervention, compared to controls. Moreover, we hypothesized that any positive changes could be stronger in circumstances where language processing is more difficult, i.e., in bilinguals compared to monolinguals, and in L2 compared to L1. Bilingualism and language dominance effects should be further characterized by differential performance in the different sentence context conditions.

Method

Participants

Older English monolinguals and older Norwegian-English bilinguals were recruited from the Birmingham (UK) and Kristiansand (Norway) communities and were tested at the University of Birmingham and the University of Agder, respectively. The eligibility criteria for monolinguals were that they were native speakers of British English, and that they would not be able to hold a simple conversation in any other language. Bilinguals completed an adaptation of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) and were eligible if: (i) Norwegian was their first acquired language; (ii) Norwegian was their dominant language and English was their second most dominant language; and (iii) they self-rated both their speaking and reading proficiency in English with at least 3 on a 0–10 scale with 0 being "none" and 10 being "perfect" (see Table 1). All participants were screened for mild cognitive impairment using the MoCA (Nasreddine et al., 2005) in its original English (monolinguals) and the Norwegian 7.1 version (bilinguals; available at www.mocatest.org); participants that scored <23 were excluded (Carson et al., 2018). All participants also reported no health conditions that would prevent safe participation in HIIT (e.g., cardiovascular, metabolic, respiratory, neurological, kidney, liver, or cancerous disease). Furthermore, participants had to: (i) obtain consent from their personal physician to participate in the

Table 1. Demographic characteristics of participants.

Characteristic	English monolinguals ($N = 78$)		Norwegian-English bilinguals ($N = 78$)	
	Exercisers ($N = 39$)	Controls ($N = 39$)	Exercisers ($N = 39$)	Controls ($N = 39$)
Mean age (*SD*)	65.69 (4.73)	65.13 (4.90)	66.97 (5.51)	66.43 (4.11)
Education				
Compulsory, n (%)	10 (25.64)	9 (23.08)	2 (5.13)	6 (15.38)
Upper secondary, n (%)	16 (41.03)	9 (23.08)	7 (17.95)	10 (25.64)
Undergraduate degree, n (%)	3 (7.69)	12 (30.76)	20 (51.28)	17 (43.59)
Postgraduate degree, n (%)	10 (25.64)	9 (23.08)	10 (25.64)	6 (15.38)
Mean MoCA score (*SD*) a	27.25 (1.98)	27.87 (1.43)	27.55 (1.61)	27.17 (1.52)
LEAP-Q [b]				
Mean self-rated proficiency in speaking L1 (SD)	—	—	9.64 (0.74)	9.31 (0.95)
Mean self-rated proficiency in reading L1 (SD)	—	—	9.87 (0.41)	9.42 (1.20)
Mean self-rated proficiency in speaking L2 (SD)	—	—	6.44 (1.68)	6.69 (1.72)
Mean self-rated proficiency in reading L2 (SD)	—	—	7.25 (1.60)	7.00 (1.74)

[a]For bilinguals, we could not track the ID of 7 questionnaires so the figures refer to 36 bilingual exercisers and 35 bilingual controls.
[b]Proficiency level based on self-ratings using a scale of 0–10 with 0 being "none" and 10 being "perfect."

study (for the bilinguals), or pass ECG and blood pressure screening overseen by a physician (for the monolinguals); (ii) self-report to complete less than 150 minutes per week of moderate-to-vigorous intensity exercise. All participants had normal or corrected-to-normal vision and hearing and gave written informed consent.

The data for this study were collected as part of a larger project that was publicly registered on OSF (see https://osf.io/d7aw2/ for a description of all measures which were part of the project). The current study reports analyses from matched (criteria described in the next paragraph) samples of 80 older bilinguals (40 exercisers (E), 40 controls (C)) and 80 older monolinguals (40 E, 40 C). For technical reasons, the data from two monolinguals were lost. We also excluded one bilingual who had response latencies <150 ms in all trials of one session, and another bilingual for whom less than 50% of trials was recorded. Thus, our analysis was conducted on the data from 78 older bilinguals (39 exercisers (25 females) and 39 controls (23 females)) and 78 monolinguals (39 exercisers (18 females) and 39 controls (20 females)).

Monolinguals and bilinguals were matched for age (64.5 ± 4.8 vs. 66.7 ± 4.8 years, respectively; t-test: t=-1.68, df = 153.99, p = 0.10). Age was also not significantly different for the C and E groups (65.78 ± 4.54 vs. 66.33 ± 5.14 years, respectively; t-test: t=-0.71, df = 151.66, p = 0.48). Education level was not significantly different for monolinguals and bilinguals, nor for exercisers and controls (cumulative link model where Education (factor with 4 levels: CompulsoryEducation, UndergraduateDegree, UpperSecondary, PostgraduateDegree; see distribution across categories in Table 1) was analyzed as function of bilingualism and group (coefficients, *SE*s and *p*-value for the factors bilingualism and group were, respectively, b=-0.43, SE = 0.41, p = 0.29; b = 0.25, SE = 0.58, p = 0.66).

Procedure

Eligible participants from each language group were randomized into an exercise or a control group, stratified by age and sex using a bespoke algorithm.[2] Participants were first assessed on CRF and then, on a different day, on language comprehension. Bilingual participants returned for a third session (≥2 days after the second session) to perform the comprehension task in L1 or L2 (the order of the tested language was counterbalanced between participants). During the following 26 weeks (i.e., 6 months), exercise group participants completed a training protocol (see below), whereas control group participants were asked to maintain their normal levels of physical activity. Post-intervention tests (following the 26-week intervention) were the same as the pre-intervention tests.

Our study was approved by the University of Birmingham's Institutional Ethics Board (ERN_20–1107), the University of Agder's Ethical Board at the Faculty of Health and Sport Science, the Norwegian Center for research data (Ref. 239577), and the Regional Committee for Medical and Healthcare Research Ethics in Norway (REK sør-østC, ref. 163931).

Training protocol

Exercisers followed a standardized 26-week home-based HIIT program involving three exercise sessions per week with required individual adjustments. Participants began with a 4-week "familiarization period," where exercise intensity and duration were lower. Following this, participants were instructed to target an intensity of >80% of peak heart rate (HR_{peak}). HR_{peak} was determined from the incremental treadmill test (detailed below). Participants completed one circuit and two interval training sessions each week, based on detailed written and video instructions. Each session lasted 40–60 minutes, including a warm-up and cool-down. Circuit training involved 3 × 45 second sets of six exercises, aiming for >80% HR_{peak}, with 30- or 90-seconds rest between sets or exercises, respectively. Interval training sessions involved alternating between 2 minutes of high-intensity exercise and active recovery. Participants were attempting to reach >80% HR_{peak} by the end of each high-intensity interval, and the majority opted to walk uphill, jog, or run. The number of intervals gradually increased throughout the program (refer to Fosstveit et al., 2024, for further details). Participants used a Polar Unite watch, Polar H9 chest heart rate monitor (Polar Electro Oy, Kempele, Finland), and a training logbook to monitor exercise, and they had monthly follow-ups with personal exercise coaches. Adherence to the exercise program was high, as measured by cumulative MET-mins which was 10,596.88 ± 4756.005, above the planned exercise load of 8567.

Cardiorespiratory fitness assessment

Cardiorespiratory fitness ($VO2_{peak,}$ mL/Kg/min) was measured using a combined modified Balke treadmill protocol (Balke & Ware, 1959). Respiratory gases (oxygen consumption and carbon dioxide production) were recorded continuously using a facemask and a - Vyntus™ CPX metabolic cart (Vyaire, Mettawa, Illinois, USA). Participants started with a self-paced walking warm-up, transitioning to a standardized progressive walking protocol until reaching volitional exhaustion. The protocol began at 3.8 km/h with a 4% incline, incrementally increasing the gradient by 3% every 4 minutes until a blood lactate concentration ([La–]b) of 2.1 mmol/L above baseline. Key metrics, including [La–]b, VO2, heart rate (HR), and ratings of perceived exertion (Borg, 1970), were recorded at the end of each 4-minute stage and posttest. After determining the lactate threshold (LT), exercise intensity was increased each minute via increased gradient (2% increments up to a maximum incline of 20%) or speed (0.5 km/h increments) until participants reached volitional exhaustion. $VO2_{peak}$ was defined as the mean of the two highest measurements recorded over 30 seconds. A detailed description of all measurements and assessment can be found in Fosstveit et al. (2024).[3] For a small subset of monolingual participants, we use estimate $VO2_{peak}$ values, following methods described in Feron et al. (2024).

Language comprehension assessment

In the word monitoring task (Marslen-Wilson & Tyler, 1975, 1980), participants had to detect target words (e.g., "spatula") in auditorily presented sentences that could be lists of words in random order (RWO; example a), low-constraint (example b), or high-constraint sentences (example c). Before each sentence, participants saw the target word, in the center of the screen, for 1000 ms (font "Consolas," size 30). Five hundred milliseconds after target word offset, the spoken sentence started playing. Participants were instructed to pay attention to the written target word and to press the spacebar on the keyboard as soon as they heard it in the spoken stimuli. Each trial ended 2000 ms after the end of the audio file. Then, the next trial was presented, starting with a central fixation cross ("+") presented for 500 ms, followed by a 1000 ms blank screen, and then the new target word. Monitoring RTs were measured from the onset of each target word in the spoken stimulus. The task took around 15 minutes to complete.

(a) Tried I find to quickly the spatula without pancake it to flip the breaking.
(b) I tried to quickly find the spatula to flip the pancake without breaking it.
(c) I flipped the pancake with the spatula without breaking it.

We created four different sets of 60 target words/sentences, two sets in British English and two sets in Norwegian Bokmål,[4] that were used at pre- and post-testing (a full description of materials is provided in Supplemental online material 1). The sentences were recorded by female native speakers of Standard British English and Norwegian Bokmål. Each target word was embedded in each of the three context conditions across three different lists (Latin Square Design). In addition to the experimental items, 12 fillers were created using different target words/sentences. Therefore, each list contained 72 trials. These were divided into four blocks of 18 items (containing 15 experimental and 3 filler items). The order of trials within a block was pseudo-randomized so that no more than three trials of the same condition occurred after each other. Two versions of each list were created with a different order of presentation of the blocks. The final experiment therefore consisted of two unique sets with six lists each, per language. Participants were pseudo-randomly assigned to lists and sets, with the constraint that they would not repeat a list/set from pre- to post-testing.

Word monitoring data pre-processing

We first excluded extreme latencies, i.e., values <150 ms and >1500 ms (827 of 28,080 trials, 2.95%). As the distributions of RTs are typically skewed (e.g., Ratcliff, 1993), we transformed the raw measures. We used the "boxcox" function in R, inputting a model of RT as a function of the variables bilingualism (mono- vs. bilingualism), context condition (RWO, low-constraint, high-constraint), group (exercisers vs. controls), time of intervention (pre- vs. post) and language (L1 vs. L2). The estimated optimal value of the λ-coefficient for the Box-Cox power transformation was −0.06 (i.e., closer to 0 than to 1 or −1), making log transformation the more appropriate for our data (Box & Cox, 1964; Kliegl et al., 2010; Venables & Ripley, 2002). In addition, we inspected normality through QQ-plots and further removed, for each participant and session (i.e., for pre- and post-testing sessions and, in addition, for bilinguals, for L1 and L2 sessions), values more than 2.5*SD* from their mean (log-transformed) times (as recommended in Howell, 2006; Ratcliff, 1993). This led to the exclusion of 455 outliers from 27,253 trials (1.67%). In total, we excluded 4.57% of the data.

Analysis

We first looked at how mono- and bilinguals' comprehension was differently affected by the exercise intervention. For that purpose, we restricted analyses to the data from L1 (i.e., excluding the data from bilinguals' L2 processing). The second set of analyses was restricted to the data from bilingual participants and compared processing in first and second language. For both the monolingual-bilingual and the L1-L2 comparisons, we first conducted group analyses to examine effects of bilingualism (or language L1-L2) and exercise intervention on language comprehension. Then, to gain insight into the mechanisms underlying these effects, we further conducted a mediation analysis to evaluate whether and how CRF ($VO2_{peak}$) mediated any effects observed at group level.

Group analysis

We analyzed (log-transformed) reaction times (log-RTs) as a function of the context conditions (low-constraint sentences, high-constraint sentences, or random words lists: RWO), the variables coding the exercise intervention (group: C vs. E, and time of intervention: pre- vs. post-), and bilingualism (for the mono-bilingual comparison: bilingual vs. monolingual) or language (for the L1-L2 comparison: L1 vs. L2). Differences between RWO and low-constraint conditions index the use of syntactic information, whereas differences between low- and high-constraint sentences indicate the use of semantic information. Accordingly, we coded the context condition using forward difference coding, where each contrast compares adjacent levels (each level minus the next level): The first contrast is "RWO" minus "low-constraint" (termed "syntax") and the second one is "low-constraint" minus "high-constraint" (termed "semantics"). The other predictors were centered (see model's output for coefficients). Finally, the analyses were adjusted for sex and age, by including these as covariates (main effects only).

We used linear-mixed effects models (LMM; Baayen et al., 2008), as implemented by *lmer* in *R*, which allow for simultaneous estimation of the variance between participants and between items (participants: 156 (78 in the L1-L2 comparison), and items (i.e., words/ sentences): 240, were random factors in the model). We fitted full models (all main effects and interactions) with a maximal-random structure justified by the design and necessary to avoid nonconvergence, which tends to increase with the complexity of the model, especially the random effects structure (Barr et al., 2013). In particular, we dropped by-item random slopes for the fixed effects bilingualism, group and time of intervention (refer to the summaries of models for its syntax), thereby simplifying the random effects structure as advised in Barr et al. (2013). We report the predictors' coefficients (β values), *SE*, *t* values, and the derived *p* significance values (by treating the *t*-statistic using the standard normal distribution as a reference; e.g., Baayen et al., 2008, footnote 1).

Mediation analysis

To gain insight into the mechanisms underlying the effects of the physical intervention on language comprehension, we further conducted mediation analysis to evaluate whether and how $VO2_{peak}$, indexing CRF, might mediate those effects. In the pre-registration of our study, we hypothesized that changes in performance between pre- and post-intervention points in time might be mediated by the degree of change in fitness level, but we did not specify an analytical approach that could be used to address this question. The choice for the current mediation analysis was therefore determined after our study was preregistered.

Mediation analysis (e.g., Baron & Kenny, 1986; MacKinnon et al., 2012) is intended to examine whether a third variable can account for the relationship between an independent variable (IV, also called treatment) and an outcome (the dependent variable, DV), therefore explaining how and why the predictor affects the outcome. This analysis decomposes the total IV-outcome effect into an indirect effect through a mediator and a direct effect that is not through the mediator. We chose as a mediator of interest a prototypical measure of CRF, $VO2_{peak}$ (measured relative to weight, i.e., mL/Kg/min), which is the maximum volume of oxygen that can be inspired at the point of voluntary physical exhaustion during maximal exercise tests.

Mediation analysis has been used in the context of health or exercise interventions (e.g., Vidoni et al., 2015, see for a review; Cashin & Lee, 2021) and can be performed using a more traditional approach (Baron & Kenny, 1986) or more recent approaches that allow to estimate the amount and significance of any mediation effects. Here, we resorted to one of these new approaches by using the R package "mediation" (Tingley et al., 2013, 2014). The process involves specifying two statistical models: (a) the "mediator model," for the conditional distribution of the mediator ($VO2_{peak}$) given the IV, and (b) the "outcome model," for the conditional distribution of the outcome (log-RTs) given the IV and the mediator ($VO2_{peak}$). The outputs of the two models are then fed into the "mediate" function, which computes the estimated average causal mediation effects (ACME) and the average direct effects (ADE).

In the case of exercise interventions with a pretest – posttest control group design, the outcome is measured at two different points in time, making the IV the combination of two variables (in our study, Group (Group: Control vs. Exercise) and time of intervention (Time: pre- vs. post-)). The current version of the "mediate" function does not allow to specify more than one IV. Therefore, we coded our IV as a non-binary treatment variable, i.e., a "condition" variable ("cond") with four categorical levels crossing Group (C vs. E) and Time of intervention (pre- vs. post-): Control_pre, Control_pos, Exercise_pre, Exercise_pos. This procedure allows to further add a moderator variable (that is, a variable for which levels the mediated effects are hypothesized to vary), i.e., bilingualism (or language, for the L1-L2 comparison). The moderator is introduced in both the mediator and outcome models, including interactions with the treatment ("cond") and mediator ($VO2_{peak}$) variables that are theoretically justified (Tingley et al., 2014; see e.g.; Montoya et al., 2023; Rijnhart et al., 2022, for alternative approaches).

Similar to the group analysis, we ran LMMs, as implemented by *lmer* in *R*, with a maximal-random structure justified by the design (although in this case by-participant, but not by-item, random intercepts were added, as the current 'mediation' function does not allow for having two random factors simultaneously; see summaries of models for specific syntax). Models were adjusted for sex and age, by including these as covariates (main effects only).

Table 2. Summary of the model for the group analysis on the effects of the exercise intervention on language comprehension in mono- and bilinguals.

(log) RT to target, listening comprehension				
Predictors	Est.	*SE*	*t*	*p*
(Intercept)	2.656	0.01	407.46	**<.01**
Syntax	**0.082**	**0.00**	**18.24**	**<.01**
Semantics	**0.036**	**0.00**	**8.24**	**<.01**
Bilingualism [bil, −0.5; mon, 0.5]	**−0.038**	**0.01**	**−2.91**	**<.01**
Time [pos, −0.5; pre, 0.5]	**0.005**	**0.00**	**3.57**	**<.01**
Group [C, −0.5; E, 0.5]	−0.015	0.01	−1.27	0.20
Sex [F, −0.45, M, 0.65]	−0.003	0.01	−0.23	0.82
Age [continuous, −7.03 to 14.97]	**0.004**	**0.00**	**3.11**	**0.00**
Syntax:Group	−0.003	0.01	−0.47	0.64
Semantics:Group	−0.001	0.00	−0.20	0.84
Bilingualism:Group	−0.009	0.02	−0.40	0.69
Time:Group	**0.009**	**0.00**	**3.06**	**<.01**
Syntax:Time	0.001	0.00	0.24	0.81
Semantics:Time	0.001	0.00	0.25	0.80
Bilingualism:Time	**0.027**	**0.00**	**9.59**	**<.01**
Syntax:Bilingualism	**0.033**	**0.01**	**3.65**	**<.01**
Semantics:Bilingualism	−0.006	0.01	−0.70	0.49
Syntax:Time:Group	−0.011	0.01	−1.57	0.12
Semantics:Time:Group	−0.001	0.01	−0.14	0.89
Bilingualism:Time:Group	**0.026**	**0.01**	**4.55**	**<.01**
Syntax:Bilingualism:Group	0.016	0.01	1.38	0.17
Semantics:Bilingualism:Group	0.005	0.01	0.61	0.54
Syntax:Bilingualism:Time	−0.002	0.01	−0.23	0.82
Semantics:Bilingualism:Time	−0.006	0.01	−0.85	0.40
Syntax:Bilingualism:Time:Group	0.005	0.01	0.32	0.75
Semantics:Bilingualism:Time:Group	−0.024	0.01	−1.73	0.08

Note: the syntax of the model is: lmer(depM ~1 + Syntax + Semantics + Bilingualism + Time + Group + Sex + Age + Syntax:Group + Semantics:Group + Bilingualism:Group + Time:Group + Syntax:Time + Semantics:Time + Bilingualism:Time + Syntax:Bilingualism + Semantics:Bilingualism + Syntax:Time:Group + Semantics:Time:Group + Bilingualism:Time:Group + Syntax:Bilingualism:Group + Semantics:Bilingualism:Group + Syntax:Bilingualism:Time + Semantics:Bilingualism:Time + Syntax:Bilingualism:Time:Group + Semantics:Bilingualism:Time:Group + (1 | subj) + (1 | item) + (0 + Syntax | subj) + (0 + Syntax | item) + (0 + Semantics | subj) + (0 + Semantics| item), data = dataset, control = lmerControl(optimizer ="Nelder_Mead")).

Results

Effects of the exercise intervention on latencies to spoken comprehension in monolinguals and bilinguals

The summary of the model for the group analysis on the mono-bilingual comparison (in L1) is presented in Table 2, and the effects are illustrated in Figure 1. The effects of sentence context (syntax and semantics, i.e., shorter detection times in low-constraint relative to random word order, and in high- compared to low-constraint) and of bilingualism (monolinguals were faster than bilinguals) are shown in Figure 1 (a). Figure 1 (a) also shows that monolinguals outperformed bilinguals in the use of syntactic information, as indicated by the bigger difference, in the first group, between latencies in RWO and normal sentences (interaction of bilingualism with syntax). No bilingualism differences were found in use of semantic information (i.e., interaction of semantics with bilingualism). Bilinguals were also slower than monolinguals, across conditions. The main effect of chronological age indicates the expected slowed processing associated with aging. These results replicate the ones reported in Fernandes et al. (2024), now in a larger data set of older adults.

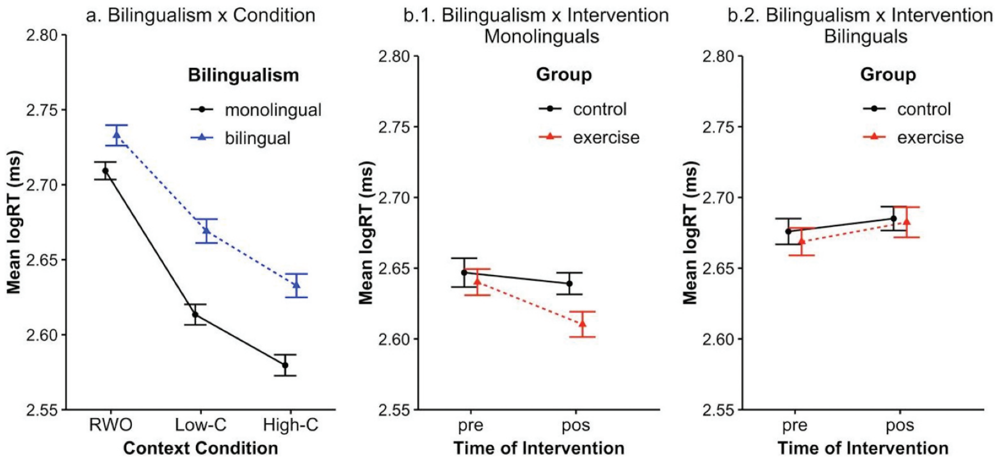

Figure 1. Mean latencies during comprehension for each bilingualism group and context condition (a; RWO, low-constraint, high-constraint), and for the monolingual (b.1) and bilingual (b.2) control and exercise groups, at pre- and post-testing. Error bars represent standard errors on means.

The interaction between group and time of intervention indicates that there was a beneficial effect (faster detection times) of the intervention for exercisers, with shorter latencies in post- compared to pre-intervention testing. However, these factors also interacted with bilingualism (Figure 1 (b)). To further investigate this interaction, we created a four-level variable combining bilingualism (bilingual vs. monolingual) with time of intervention (pre- vs. post-intervention), across context conditions. This four-level variable was contrast coded so that we had three contrasts: bilingual_pre-intervention × bilingual_post-intervention; monolingual_pre-intervention × monolingual-_post-intervention, and bilingual × monolingual. We ran an LMM adding also group (C vs. E), age, and sex. The contrast bilingual × monolingual was significant ($b = -0.039$, $SE = 0.01$, $p < 0.01$), indicating a bilingual disadvantage. The contrasts bilingual_pre-intervention × bilingual_post-intervention and monolingual_pre-intervention × monolingual_post-intervention were both significant, but with opposite signs ($b = 0.009$, $SE = 0.01$, $p < .01$ and $b = -0.018$, $SE = 0.01$, $p < .01$, respectively): Monolinguals were faster, whereas bilinguals were slower, in post- compared to pre-testing. In the first case only, the contrast interacted with group (monolingual_pre-intervention × monolingual_post-intervention × group; $b = -0.022$, $SE = 0.01$, $p = <.01$), indicating that only monolinguals in the exercise group were faster after the intervention, but that bilinguals were slower across groups.

For the mediation analysis, since there were no interactions between sentence context and the conditions coding the exercise intervention, we did not include sentence context as a predictor in the models. Table 3 presents the "mediator model" regressing the mediator ($VO2_{peak}$) against the exercise intervention. As mentioned in the Analysis section, the intervention was coded through a 4-level categorical variable combining the groups and the time of intervention: Control_pre, Control_pos, Exercise_pre, Exercise_pos. We chose the exercise group pre-intervention (Exercise_pre) as the reference level of this variable, as the contrast of greatest interest was the one between participants in the exercise

Table 3. Summary of the mediator model regressing VO2peak on the independent variable coding the physical exercise intervention and on the moderator (Bilingualism).

Mediator model: VO2peak (Ref. Level = Exercise_pre)				
Predictors	Est.	*SE*	*t*	*p*
(Intercept)	**27.917**	**0.50**	**55.62**	**<.01**
cond Control_pos	−0.488	0.70	−0.69	0.49
cond Exercise_pos	**1.668**	**0.03**	**65.43**	**<.01**
cond Control_pre	−0.246	0.70	−0.35	0.73
Bilingualism	0.157	1.02	0.15	0.88
Age [−7.03 to 14.97]	**−0.173**	**0.07**	**−2.34**	**0.02**
Sex [F, −0.45; M, 0.55]	**4.073**	**0.71**	**5.72**	**<.01**
cond Control_pos:Bilingualism	−2.045	1.41	−1.45	0.15
cond Exercise_pos:Bilingualism	**−0.265**	**0.05**	**−5.20**	**<.01**
cond Control_pre:Bilingualism	−1.538	1.41	−1.09	0.28

Note: the syntax of the model is: lmer(depM ~ cond + Bilingualism + cond:Bilingualism+ Age + Sex + (1 |subj), data=dataset).

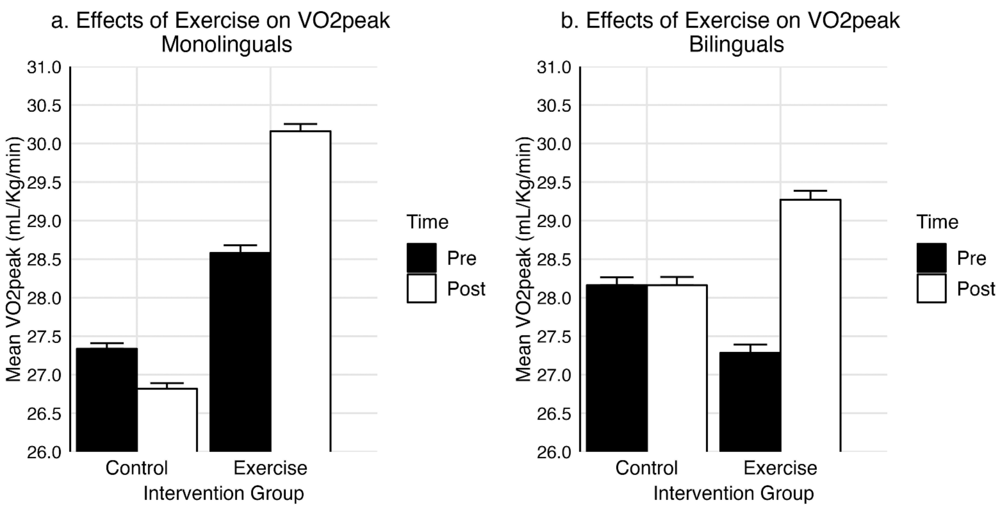

Figure 2. Mean $VO2_{peak}$ for monolinguals (a) and bilinguals (b) on the control and exercise groups, at pre- and post-testing. Error bars represent standard errors on means.

group at the moment of post-intervention relative to this reference level. Bilingualism (centered) and its interactions with the condition levels were also entered in the model. Finally, the model included the covariates age and sex (main effects only).

The effects of the intervention on the CRF measure are plotted in Figure 2. The effect of "cond Exercise_pos" indicates that the intervention was successful in inducing higher $VO2_{peak}$, which increased from pre- to post-testing in the exercise group. This effect interacted with bilingualism. Simple effects' analyses on each group showed that the increase in $VO2_{peak}$ from pre- to post-testing in exercisers was smaller in monolinguals than in bilinguals ($b = 1.54$, $SE = 0.04$, $p < .001$ and $b = 1.80$, $SE = 0.03$, $p < .001$, respectively). For the monolingual group, there was a trend for a decrease in $VO2_{peak}$ from pre- to post-testing in the control group (b = -1.58, SE = 0.85, p = .06).

Table 4. Summary of the outcome model regressing (log-)RTs on the variable coding the physical exercise intervention, the mediator (VO2$_{peak}$), and the moderator (bilingualism).

Outcome Model: (log)RT (Ref. Level=Exercise_pre)

Predictors	Est.	*SE*	*t*	*p*
(Intercept)	2.651	0.01	281.74	**<.01**
cond Exercise_pos	0.002	0.00	0.75	0.45
cond Control_pre	0.003	0.01	0.21	0.83
cond Control_pos	0.002	0.01	0.15	0.88
Bilingualism [bil, −0.5; mon, 0.5]	−0.027	0.02	−1.43	0.15
VO2peak [−11.45 to 12.90]	**−0.005**	**0.00**	**−4.27**	**<.01**
Age [−7.03 to 14.97]	**0.003**	**0.00**	**2.23**	**0.03**
Sex [F, −0.45; M, 0.55]	0.009	0.01	0.69	0.49
cond Exercise_pos:VO2peak	−0.001	0.00	−1.04	0.30
cond Control_pre:VO2peak	−0.001	0.00	−0.49	0.62
cond Control_pos:VO2peak	−0.003	0.00	−1.84	0.07
cond Exercise_pos:Bilingualism	**−0.025**	**0.01**	**−3.83**	**<.01**
cond Control_pre:Bilingualism	−0.016	0.03	−0.60	0.55
cond Control_pos:Bilingualism	−0.035	0.03	−1.34	0.18
Bilingualism:VO2peak	**−0.005**	**0.00**	**−2.42**	**0.02**
cond Exercise_pos:Bilingualism:VO2peak	−0.001	0.00	−1.19	0.23
cond Control_pre:Bilingualism:VO2peak	0.002	0.00	0.50	0.62
cond Control_pos:Bilingualism:VO2peak	0.000	0.00	0.01	1.00

Note: the syntax of the model is lmer(depM ~ cond + Bilingualism + VO2 + cond:VO2 + cond:Bilingualism + Bilingualism:VO2 +cond:Bilingualism:VO2 +Age + Sex + (1 |subj), data=dataset, control = lmerControl (optimizer ="Nelder_Mead")).

The summary of the "outcome model," regressing (log-)RT on the intervention (Control_pre, Control_pos, Exercise_pre, Exercise_pos), the mediator (VO2$_{peak}$), the moderator (bilingualism), and the covariates age and sex is presented in Table 4.

The difference between the performance of the exercisers before and after the intervention was modulated by bilingualism (cond Exercise_pos:Bilingualism). Follow-up simple effects' analyses for each bilingual group separately, with the predictors "condition" and VO2$_{peak}$ (and age and sex as covariates) showed that monolinguals in the exercise group were faster at post- compared to pre-testing (marginally significant effect of cond Exercise_pos: b = -0.009, SE = 0.01, p = .06). In contrast, bilinguals in the exercise group were slower at post- compared to pre-testing (cond Exercise_pos: b = 0.014, SE = 0.01, p < .001).[5] There was a main effect of VO2$_{peak}$ in the monolingual group (b = -0.007, SE = 0.01, p < .001), but not in the bilingual group (b = -0.002, SE = 0.01, p = .24), suggesting that the two groups were differently affected by changes in CRF. This is illustrated in Figure 3, plotting changes in comprehension as a function of changes in VO2$_{peak}$, computed for each participant (note that these change scores were computed and used for the purpose of an easier visualization only, but individual observations, and not change scores, were used in all analyses).

To measure the mediation effects of VO2$_{peak}$ we fed the mediator and the outcome models into the "mediate" function, separately for the levels of the moderator (bilingualism), and for the specific contrasts between the levels of the condition variable coding the intervention.[6] The average mediated (ACME) and direct (ADE) effects are shown in Table 5. The estimates for the mediation effects, direct effects, and proportion of total effect mediated correspond to the levels of the condition variable, which are contrasted in pairs (e.g., contrast between pre- and post-testing for the monolinguals in the Exercise group).

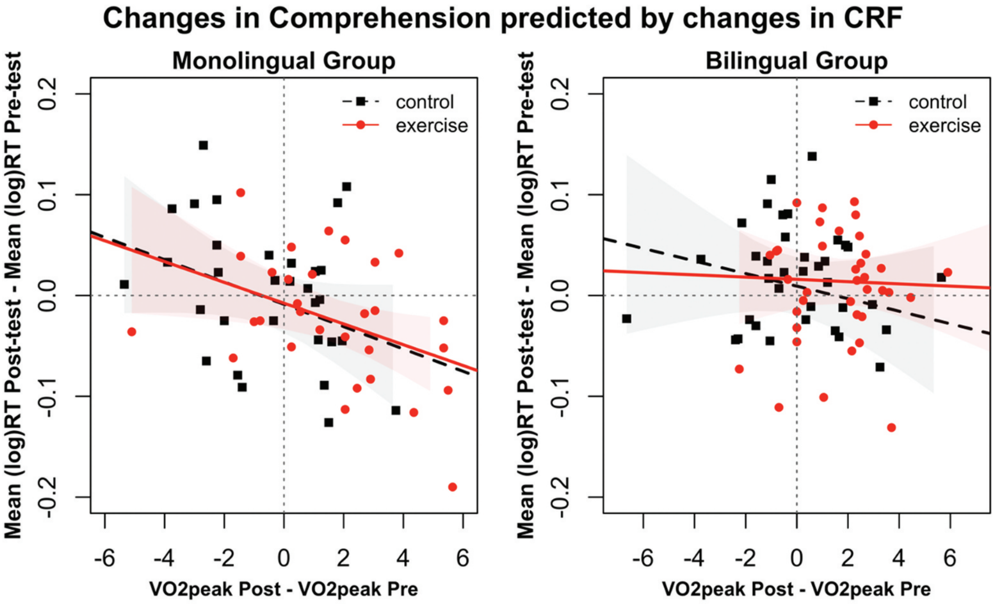

Figure 3. Observed changes in comprehension performance as a function of changes in VO2peak from pre- to post-testing, for the monolinguals (left panel) and bilinguals (right panel) and the control (black squares and dashed lines) and the exercise (red circles and solid lines) groups. On each panel, the bottom right quadrant corresponds to increases in VO2peak and decreases in RT to comprehension.

For monolinguals, all mediated effects were significant. The proportion of the total effect of the intervention that was mediated through $VO2_{peak}$ is estimated to be 0.20 in the exercise group, and −0.05 in the control group. This suggests that the intervention increased the value of $VO2_{peak,}$ which in turn decreased RTs during listening comprehension for exercisers. The direct effect was significant in the control group, but only marginally significant in the exercise group. This indicates that the intervention also led to faster performance of controls directly (direct effect, i.e., just the passage of time caused by the intervention, in the control group), or through different causal paths (i.e., attributable to changes in some parameter that was not measured but not through changes in $VO2_{peak}$).

In the bilingual group, none of the effects were significant. This seems to suggest that the slowed language comprehension by bilinguals after, compared to before the intervention, across groups, cannot be explained by either direct or (CRF) mediated effects of the physical exercise intervention.

Effects of the exercise intervention on latencies to spoken comprehension in L1 and L2

In the group analysis (Table 6) we examined how bilinguals' sentence comprehension was affected, in L1 and L2, by the exercise intervention, coded through the two variables group (Group: C vs. E) and time of intervention (time: pre- vs. post). As before, effects of sentence context were examined using forward difference coding and two contrasts:

Table 5. Causal mediation analysis, quasi-Bayesian confidence intervals, 1000 simulations.

	Monolinguals control ="Exercise_pre," treated = "Exercise_pos"				Monolinguals control ="Control_pre," treated = "Control_pos"			
	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value
ACME (control)	**−0.011**	**−0.02**	**−0.01**	**<.01**	**0.004**	**0.01**	**0.01**	**<.01**
ACME (treated)	**−0.013**	**−0.02**	**−0.01**	**<.01**	**0.005**	**0.00**	**0.01**	**<.01**
ADE (control)	−0.046	−0.09	0.00	0.07	**−0.087**	**−0.16**	**−0.02**	**0.02**
ADE (treated)	−0.048	−0.10	0.00	0.07	**−0.086**	**−0.15**	**−0.02**	**0.02**
Total Effect	**−0.058**	**−0.11**	**−0.01**	**0.02**	−0.082	−0.15	−0.02	0.02
Prop. Mediated (control)	**0.182**	**0.07**	**0.97**	**0.02**	**−0.043**	**−0.18**	**−0.01**	**0.02**
Prop. Mediated (treated)	**0.216**	**0.11**	**0.94**	**0.02**	**−0.060**	**−0.20**	**−0.03**	**0.02**
ACME (average)	**−0.012**	**−0.02**	**−0.01**	**<.01**	**0.004**	**0.00**	**0.01**	**<.01**
ADE (average)	−0.047	−0.10	0.00	0.07	**−0.086**	**−0.16**	**−0.02**	**0.02**
Prop. Mediated (average)	**0.199**	**0.09**	**0.95**	**0.02**	**−0.052**	**−0.19**	**−0.02**	**0.02**
	Bilinguals control ="Exercise_pre," treated = "Exercise_pos"				Bilinguals control ="Control_pre," treated = "Control_pos"			
	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value
ACME (control)	−0.004	−0.01	0.00	0.17	0.000	0.00	0.00	0.80
ACME (treated)	−0.004	−0.01	0.00	0.16	0.000	0.00	0.00	0.80
ADE (control)	0.016	−0.02	0.06	0.45	−0.027	−0.07	0.01	0.18
ADE (treated)	0.016	−0.03	0.06	0.48	−0.027	−0.07	0.01	0.18
Total Effect	0.012	−0.03	0.05	0.57	−0.027	−0.07	0.01	0.18
Prop. Mediated (control)	−0.131	−3.13	1.75	0.57	0.001	−0.04	0.04	0.81
Prop. Mediated (treated)	−0.092	−3.64	2.17	0.64	0.001	−0.04	0.05	0.81
ACME (average)	−0.004	−0.01	0.00	0.15	0.000	0.00	0.00	0.80
ADE (average)	0.016	−0.02	0.06	0.46	−0.027	−0.07	0.01	0.18
Prop. Mediated (average)	−0.111	−3.43	1.95	0.60	0.001	−0.04	0.05	0.81

Note: ACME are the average causal mediation effects, and ADE are the average direct effects. Prop. Mediated is the proportion of the effect that is mediated.

"RWO" minus "low-constraint" (i.e., "syntax") and the "low-constraint" minus 'high-constraint" (i.e., "semantics"). All other variables were centered (see Table 6 for coefficients).

As expected, participants were slower in L2 compared to L1 (Figure 4 (a)). The effects of context ("syntax" and "semantics") confirm that listeners made use of both syntactic and semantic information during comprehension. However, the effect of "semantics" was modulated by language. Follow-up simple effects analyses showed that the processing advantage associated with semantics was bigger in L1 than in L2 ($b = 0.039$, $SE = 0.01$, $p < .001$ and $b = 0.022$, $SE = 0.01$, $p < .001$, respectively).

Relative to the intervention (Figure 4 (b)), we found interactions between time and group, and also between time and language. To further assess the effects, we ran an analysis separately for each language and, in each case, we created a four-level variable combining group with time of intervention. This four-level variable was contrast coded so that we had three contrasts: Control_pre × Control_pos, Exercise_pre × Exercise_pos and Control × Exercise. We also added age and sex (as main effects only). The analyses showed that bilinguals in both the control and exercise groups were slower in post- compared to

Table 6. Summary of the model for the group analysis on the effects of the exercise intervention on language comprehension in L1 and L2.

(log) RT to target, listening comprehension				
Predictors	Est.	*SE*	*t*	*p*
(Intercept)	2.702	0.01	318.04	**<.01**
Syntax	**0.067**	**0.00**	**14.82**	**<.01**
Semantics	**0.030**	**0.00**	**6.91**	**<.01**
Language [L1, −0.49; L2, 0.51]	**0.049**	**0.01**	**7.87**	**<.01**
Time [pos, −0.5; pre, 0.5]	**−0.006**	**0.00**	**−4.13**	**<.01**
Group [C, −0.5; E, 0.5]	−0.012	0.02	−0.74	0.46
Sex [F, −0.38; M, 0.62]	−0.028	0.02	−1.69	0.09
Age [−7.67 to 13.32]	**0.005**	**0.00**	**3.04**	**<.01**
Syntax:Group	−0.011	0.01	−1.75	0.08
Semantics:Group	0.000	0.00	−0.11	0.91
Language:Group	0.002	0.00	0.59	0.56
Time:Group	**−0.009**	**0.00**	**−3.03**	**<.01**
Syntax:Time	0.003	0.00	0.95	0.34
Semantics:Time	0.002	0.00	0.57	0.57
Language:Time	**0.005**	**0.00**	**1.95**	**0.05**
Syntax:Language	0.001	0.01	0.20	0.85
Semantics:Language	**−0.018**	**0.01**	**−2.14**	**0.03**
Syntax:Time:Group	−0.008	0.01	−1.20	0.23
Semantics:Time:Group	0.005	0.01	0.67	0.50
Language:Time:Group	−0.008	0.01	−1.41	0.16
Syntax:Language:Group	0.001	0.01	0.14	0.89
Semantics:Language:Group	0.006	0.01	0.92	0.36
Syntax:Language:Time	0.004	0.01	0.54	0.59
Semantics:Language:Time	−0.004	0.01	−0.56	0.58
Syntax:Language:Time:Group	0.009	0.01	0.68	0.50
Semantics:Language:Time:Group	−0.013	0.01	−0.94	0.35

Note: the syntax of the model is: lmer(depM ~1 + Syntax + Semantics + Language + Time + Group + Sex + Age + Syntax:Group + Semantics:Group + Language:Group + Time:Group + Syntax:Time + Semantics:Time + Language:Time + Syntax:Language + Semantics:Language +Syntax:Time:Group + Semantics:Time:Group + Language:Time:Group + Syntax:Language:Group + Semantics:Language:Group + Syntax:Language:Time + Semantics:Language:Time +Syntax:Language:Time:Group + Semantics:Language:Time:Group + (1 | subj) + (1 | item) + (0 + Syntax | subj) + (0 + Syntax | item) + (0 + Semantics | subj) + (0 + Semantics| item), data = dataset, control = lmerControl(optimizer ="Nelder_Mead")).

pre-testing when comprehending L1 ($b = 0.010$, $SE = 0.01$, $p < .001$ and $b = 0.009$, $SE = 0.01$, $p=.01$). For L2 comprehension, the slower processing in post-testing was only observed for the exercise group ($b = 0.008$, $SE = 0.01$, $p=.01$).

The mediation analysis further tested whether the effects of the exercise intervention could be, at least in part, explained by changes in $VO2_{peak}$. As in the comparison between mono- and bilinguals, we did not include sentence context as a predictor, and the intervention was coded through a 4-level categorical variable combining the groups and the time of intervention (Control_pre, Control_pos, Exercise_pre, Exercise_pos). The mediator model, showing effects of the intervention in CRF ($VO2_{peak}$), is presented in Table 7. As we have already shown (comparison mono- to bilinguals, see Figure 2 (b)), there was an increase in $VO2_{peak}$ from pre- to post-testing in the exercise group. No other effects or interactions were significant.[7]

The outcome model, assessing the effects of the exercise intervention, language, and $VO2_{peak}$ on listening comprehension, is presented in Table 8. The intervention interacted not only with language but also with $VO2_{peak}$. We conducted follow-up analyses for each language separately. For L1, the only significant effect was of the condition "cond

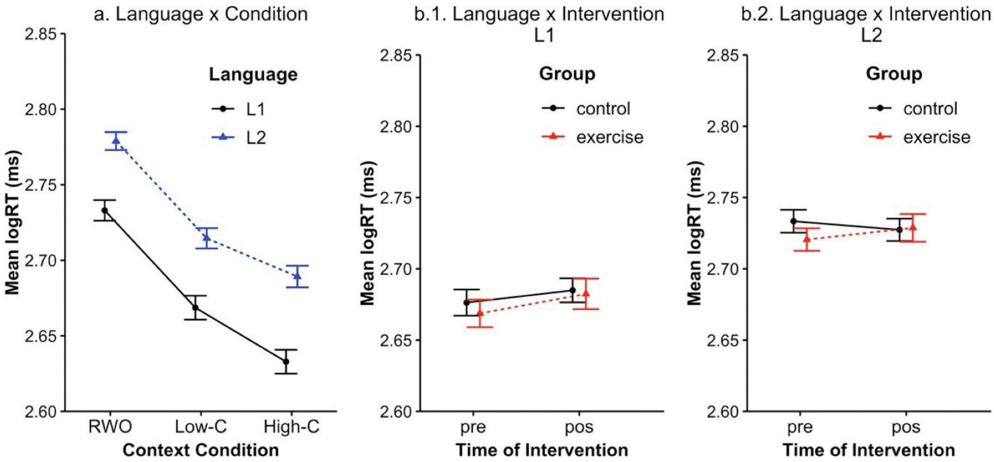

Figure 4. Mean latencies during comprehension for each language and context condition (a; RWO, low-constraint, high-constraint), and for the control and exercise groups, at pre- and post-testing, in L1 (b.1) and L2 (b.2). Error bars represent standard errors on means.

Table 7. Summary of the mediator model regressing $VO2_{peak}$ on the independent variable coding the physical exercise intervention and on the moderator (language).

Mediator model: VO2peak (Ref. Level = Exercise_pre)				
Predictors	Est.	*SE*	*t*	*p*
(Intercept)	27.458	0.75	36.78	<1e-04
cond Control_pos	0.538	1.06	0.51	0.61
cond Exercise_pos	**1.786**	**0.02**	**76.82**	**<.01**
cond Control_pre	0.534	1.06	0.50	0.61
Language	0.008	0.03	0.24	0.81
Age	−0.096	0.11	−0.87	0.39
Sex	**4.556**	**1.09**	**4.17**	**<.01**
cond Control_pos:Language	−0.017	0.05	−0.37	0.71
cond Exercise_pos:Language	−0.023	0.05	−0.49	0.62
cond Control_pre:Language	−0.004	0.05	−0.08	0.94

Note: the syntax of the model is: med.fit=lmer(depM ~ cond + Language + cond:Language + Age + Sex + (1 |subj), data=dataset).

Exercise_pos" ($b = 0.014$, $SE = 0.01$, p=.01), indicating that exercisers were slower at post-testing, compared to pre-testing (as already shown in the mono-bilingual comparison). There was no main effect of, nor interactions with $VO2_{peak}$. For L2, the effect of the condition "cond Exercise_pos" was nearly significant ($b = 0.008$, $SE = 0.01$, p = .006), but it was modulated by $VO2_{peak}$ (cond Exercise_pos:VO2: b = -0.003, $SE = 0.01$, $p < .001$; Figure 5 right panel), which indicates slower comprehension as a function of higher $VO2_{peak}$ values. However, the mediation analysis did not show any mediation effects.

Table 9 shows the output of the "mediate" function, separately for the levels of the moderator (Language), and for the specific contrasts between the levels of the condition variable coding the intervention. As reported for the mono-bilingual comparison, there were no mediation (ACME) nor direct effects (ADE) in L1 processing. In contrast, we found

Table 8. Summary of the outcome model regressing (log-)RTs on the variable coding the physical exercise intervention, the mediator ($VO2_{peak}$), and the moderator (language).

Outcome model: (log)RT (Ref. Level = Exercise_pre)				
Predictors	Est.	*SE*	*t*	*p*
(Intercept)	**2.693**	**0.01**	**231.88**	**<.01**
cond Control_pos	0.016	0.02	0.99	0.32
cond Exercise_pos	**0.010**	**0.00**	**3.37**	**<.01**
cond Control_pre	0.014	0.02	0.84	0.40
Language [L1, −0.49; L2, 0.51]	**0.053**	**0.00**	**15.31**	**<.01**
VO2peak [−11.43 to 12.92]	0.000	0.00	0.38	0.70
Age [−7.67 to 13.32]	**0.005**	**0.00**	**2.74**	**0.01**
Sex [F, −0.38; M, 0.62]	−0.020	0.02	−1.18	0.24
cond Control_pos:VO2	**−0.004**	**0.00**	**−2.55**	**0.01**
cond Exercise_pos:VO2	**−0.001**	**0.00**	**−2.97**	**0.00**
cond Control_pre:VO2	−0.003	0.00	−1.80	0.07
cond Control_pos:Language	**−0.010**	**0.00**	**−2.05**	**0.04**
cond Exercise_pos:Language	−0.006	0.00	−1.24	0.22
cond Control_pre:Language	0.003	0.00	0.68	0.49
Language:VO2	**0.002**	**0.00**	**3.60**	**<.01**
cond Control_pos:Language:VO2	−0.001	0.00	−1.11	0.27
cond Exercise_pos:Language:VO2	**−0.002**	**0.00**	**−2.61**	**0.01**
cond Control_pre:Language:VO2	−0.001	0.00	−1.40	0.16

The syntax of the model is: lmer(depM ~ cond + Language + VO2 + cond:VO2 + cond:Language + Language:VO2 + cond:Language:VO2 + Age + Sex + (1 |subj), data = dataset,control = lmerControl(optimizer ="Nelder_Mead")).

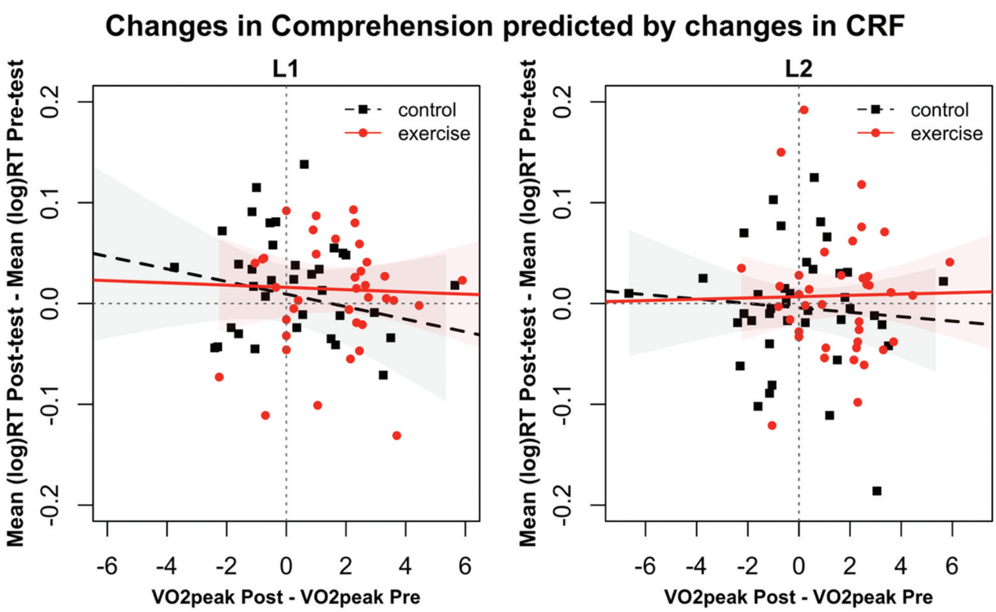

Figure 5. Observed changes in comprehension performance as a function of changes in $VO2_{peak}$ from pre- to post-testing, for L1 (left panel) and L2 (right panel) and the control (black squares and dashed lines) and exercise (red circles and solid lines) groups. On each panel, the bottom right quadrant corresponds to increases in $VO2_{peak}$ and decreases in RT to comprehension.

Table 9. Causal mediation analysis, Quasi-Bayesian confidence intervals, 1000 simulations.

	L1 control ="Exercise_pre," treated = "Exercise_pos"				L1 control ="Control_pre," treated = "Control_pos"			
	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value
ACME (control)	−0.001	0.00	0.00	0.75	0.003	0.00	0.01	0.27
ACME (treated)	−0.001	0.00	0.00	0.68	0.005	0.00	0.01	0.26
ADE (control)	0.007	−0.03	0.04	0.71	−0.027	−0.06	0.01	0.18
ADE (treated)	0.007	−0.03	0.04	0.72	−0.025	−0.06	0.01	0.19
Total Effect	0.006	−0.03	0.04	0.74	−0.022	−0.06	0.02	0.27
Prop. Mediated (control)	−0.008	−0.88	0.70	0.84	−0.074	−1.42	2.42	0.51
Prop. Mediated (treated)	−0.001	−0.98	1.03	0.96	−0.137	−1.74	2.72	0.50
ACME (average)	−0.001	0.00	0.00	0.72	0.004	0.00	0.01	0.26
ADE (average)	0.007	−0.03	0.04	0.71	−0.026	−0.06	0.01	0.18
Prop. Mediated (average)	−0.004	−0.84	0.86	0.92	−0.106	−1.58	2.56	0.50
	L2 control ="Exercise_pre," treated = "Exercise_pos"				L2 control ="Control_pre," treated = "Control_pos"			
	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value	Estimate 95%	95% *CI* Lower	95% *CI* Upper	*p*-value
ACME (control)	0.001	0.00	0.01	0.68	0.002	0.00	0.01	0.28
ACME (treated)	−0.001	0.00	0.00	0.75	0.003	0.00	0.01	0.21
ADE (control)	**−0.065**	**−0.10**	**−0.03**	**<.01**	−0.032	−0.07	0.01	0.12
ADE (treated)	**−0.067**	**−0.10**	**−0.03**	**<.01**	−0.031	−0.07	0.01	0.13
Total Effect	**−0.066**	**−0.10**	**−0.03**	**<.01**	−0.029	−0.07	0.01	0.17
Prop. Mediated (control)	−0.009	−0.09	0.04	0.68	−0.043	−0.75	0.95	0.42
Prop. Mediated (treated)	0.004	−0.04	0.07	0.75	−0.084	−0.85	0.92	0.34
ACME (average)	0.000	0.00	0.00	0.94	0.003	0.00	0.01	0.23
ADE (average)	**−0.066**	**−0.10**	**−0.03**	**<.01**	−0.032	−0.07	0.01	0.13
Prop. Mediated (average)	−0.003	−0.05	0.03	0.94	−0.064	−0.82	0.94	0.36

ACME are the average causal mediation effects, and ADE are the average direct effects. Prop. Mediated is the proportion of the effect that is mediated.

a direct effect of the intervention in L2 processing, although no mediation effects. Note that this direct effect reflects the difference between the slowing from pre- to pre-testing that was observed, for L2, in the exercise group. As the same slowing was observed for L1 across control and exercise groups, no (direct or mediated) effects of the intervention can be observed.

Discussion

Our main objectives were to investigate the potential benefits of exercise training for language comprehension in healthy older adults and how these might be affected by bilingualism and language dominance. Below, we first discuss those effects, which were observed across sentence context conditions, that is, over and above differences in syntax and semantics' processing. We then turn to a discussion of our findings on the effects of bilingualism and language dominance in language comprehension.

Effects of the exercise intervention on comprehension in monolinguals and bilinguals

The exercise intervention was beneficial for monolinguals' language comprehension. As predicted, exercisers, but not controls, were faster in detecting words in spoken sentences following the intervention, and the effect was, at least partially, due to increased CRF ($VO2_{peak}$).

This result extends previous findings in two important ways. Firstly, it extends the evidence of a beneficial effect of exercise on cognition in older age, from broader cognitive function (e.g., executive function, see Voss & Jain, 2022, for a review) to language function. Exercise has been claimed to counteract cognitive decline associated with healthy aging (see Etnier et al., 2006; Falck et al., 2019; Kramer & Erickson, 2007, for reviews), but an association between aerobic fitness and improved language processing has been demonstrated only through a positive relationship between CRF and language production (ToT states; Segaert et al., 2018). Language function is often thought to be relatively well preserved in healthy aging (e.g., Abrams & Farrell, 2011; Campbell et al., 2016), but actually there is substantial evidence for at least some age-related decline, such as the increased experience of tip-of-the-tongue states (i.e., in production; Ossher et al., 2013), and slowed and less accurate comprehension (Caplan et al., 2011; Federmeier et al., 2002, 2003; Poulisse et al., 2019; see; Abrams & Farrell, 2011; Peelle, 2018, for reviews). Here we show, for the first time, that exercise training has a beneficial effect on healthy older adults' language comprehension.

Second, we extend Segaert et al. (2018) cross-sectional evidence to an exercise intervention. Whereas no causal conclusions could be drawn previously, our randomized controlled trial allows us to infer a causal link between exercise and improved language abilities. Moreover, using mediation analysis, we showed that $VO2_{peak}$ mediated the effects of exercise training on language comprehension. This is consistent with evidence for cardiorespiratory fitness being predictive of cognitive benefits (Brown et al., 2021; Vidoni et al., 2015), and further informs about the aspects of CRF that slow cognitive aging.

Yet, this does not mean that $VO2_{peak}$ can explain the whole pattern of results, as the proportion of mediation was estimated to be only 0.2. CRF is a complex measure of physiological health that, despite being traditionally measured by $VO2_{peak}$ (or $VO2_{max}$), involves changes in other systems and metrics that are usually not accounted for in exercise interventions, such as changes associated with the neuromuscular, metabolic, and vascular systems. It is not yet completely understood how increased CRF affects the brain and, in turn, cognitive performance (Voss & Jain, 2022). CRF has been linked to increases in brain volume and white matter integrity in the frontal and temporal lobes in older adults (Burzynska et al., 2014; Colcombe et al., 2006; Voss et al., 2013). There is also evidence for effects of CRF on functional brain health, observed in cognitive performance. For example, Colcombe et al. (2004) found that aerobically trained participants showed greater activity in the prefrontal and parietal cortices, involved in spatial selection and inhibitory function while performing a Flanker task (see Rosano et al., 2010, for related evidence for processing speed). Using fMRI and a commonality analysis approach, Rahman et al. (2024) uncovered that functional activation of the brain's language networks associated with tip-of-the-tongue states is in part determined by older adults' cardiorespiratory fitness levels. However, in order to further explore the CRF hypothesis, more studies are needed that examine the relationship between exercise-induced structural and functional brain changes and cognitive (and linguistic) processing (Intzandt et al., 2021; Kramer & Erickson, 2007).

On the other hand, contrary to our prediction, the exercise intervention did not affect bilinguals' language comprehension, despite similar increases in CRF to monolinguals. In addition, for the bilinguals, $VO2_{peak}$ did not mediate any changes in language performance, and we found that both bilingual exercisers and controls were slower to detect words in spoken L1 sentences following the intervention, which is a surprising result. We might think that overall age-related slowing might be an explanation, but it seems unlikely that such changes occurred in a short 6-month period, and differentially for the age- and education-matched populations of mono- and bilinguals. We would more likely expect that participants (both exercisers and controls) would be faster following the intervention period, as a consequence of practice with the task. We found such practice effects, for both monolinguals and bilinguals, in the performance of a processing speed (letter comparison) task that all participants completed in another testing session, ruling out the explanation of a general slowing from pre- to post-testing (see Supplemental online material 2 for details on the task and the analysis).

One hallmark of bilingual language processing is the constant need to inhibit the non-target language and thus avoid cross-linguistic interference. Bilinguals are thought to select and use the intended language by using control mechanisms (Green, 1998) relying on domain-general executive functions, which explains their advantage over monolinguals on cognitive tasks involving attentional control (e.g., Bialystok et al., 2005, 2007). However, it also means that there are structural changes in the bilingual brain, compared to monolinguals, and that bilingual language processing recruits different and/or additional resources (e.g., Sulpizio et al., 2020). We speculate that such additional and differential language processing due to bilingualism might offset potential benefits of exercise or CRF.

Effects of the exercise intervention on comprehension in L1 and L2

In L2, bilingual exercisers, but not controls, were slower to detect words in spoken sentences, at post- compared to pre-testing. This would suggest that, despite the (positive) changes in $VO2_{peak}$ caused by exercise training, bilinguals' (L2) comprehension was negatively affected by the intervention, contrary to monolinguals. Yet, the effect was not mediated through CRF, meaning that the slowing is not attributable to changes in $VO2_{peak.}$ Moreover, as mentioned above, in L1 the slowing from pre- to post-testing was observed also in the control group, raising doubt about whether the effect in L2 is indeed an effect of the exercise intervention, or if it could be the case that there was a slowing from pre- to post-testing that we failed to detect in the control group's performance in L2. In both cases, the question arises as to what would cause slowed processing at post-testing.

As mentioned already, the costlier language processing in bilinguals is associated with different patterns of neural activation and more individual variation in activation patterns, especially in L2 (Perani et al., 1996; Vingerhoets, 2003). Moreover, brain activation patterns in bilinguals were found to be a function not only of L2 age of acquisition (Saur et al.,

2009) but mainly of L2 attained proficiency (Perani, 1998), and L2 proficiency is known to have a relationship with language processing cost, as the higher the proficiency the more automatic and less demanding processing is (e.g., Kroll & Stewart, 1994; Pivneva et al., 2012). Therefore, if the increased/differential language processing load for bilinguals might override potential beneficial effects of exercise, then the lack of exercise benefits (or the exercise detrimental effects) might be a function of decreased L2 proficiency.

To further test this idea, we conducted an exploratory analysis on whether and how the bilinguals' performance was modulated by L2 proficiency. One of the eligibility criteria for bilinguals in our study was that they scored their L2 reading and speaking at least 3 on a 0–10 scale (see Participants section). This was intended to avoid recruitment of very low L2 proficient speakers that might not be able to perform the L2 language tasks, but also to ensure that our sample would still include large variability in L2 proficiency, as self-rated in the LEAP-Q questionnaire. Such variability was also confirmed in a reading comprehension task that participants completed, before their pre-testing sessions, as an objective assessment of L2 proficiency. In this task participants (self-paced) read sentences of different syntactic complexity followed by a comprehension question and we computed, as the measure of proficiency, the average reading time across all sentence types (see Supplemental online material 3 for details). We performed a median split to categorize participants into two groups with low and high L2 proficiency using the reading time data. Mean scored L2 proficiency was 6877.23 (SD = 1122.36) and 10,210.58 (SD = 1672.07) in the high- and low-proficiency groups, respectively.

We conducted the same group analysis reported before for the comparison between L1 and L2 for each proficiency group separately (across sentence type conditions). We found that the pattern observed in our initial group analysis (see Figure 4 and Table 6) is mainly observed in the behavior of low L2-proficiency bilinguals. As illustrated in Figure 6, less proficient speakers were slower at post- compared to pre-testing, across languages and groups. In contrast, high proficient speakers were faster at post- compared to pre-testing, in both L1 and L2 (although the differences were only significant in the control

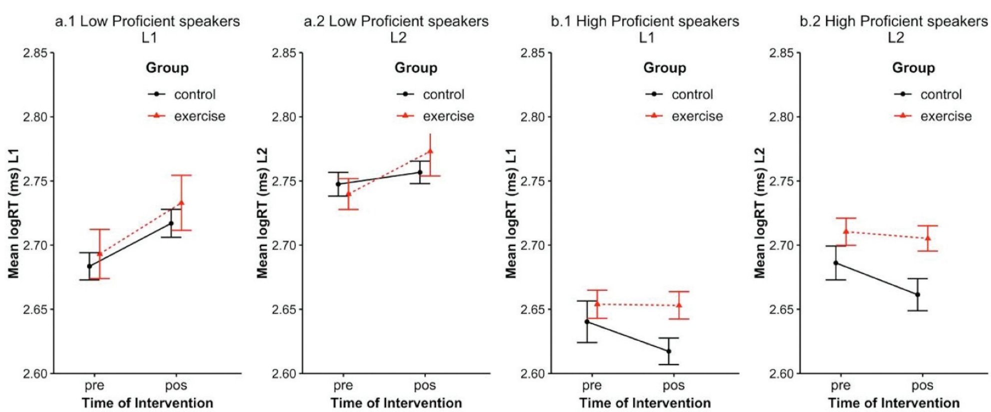

Figure 6. Mean latencies during comprehension for each language and for the control and exercise groups, at pre- and post-testing, for L2 low- (a.1, a.2) and high-proficiency (b.1, b.2) bilinguals. Error bars represent standard errors on means.

group, despite a trend for an effect in the exercise group in L2 processing; see Supplemental online material 4 for output of the analyses).

This finding suggests that the pattern of results we reported for bilinguals is mainly attributable to participants with low L2 proficiency. These participants were the ones that were slower after the 6-months intervention period. Such slowing occurred in both L1 and L2, and for the control and exercise groups.[8] In this case, the detrimental effect of time of intervention cannot be interpreted as an effect of exercise training. In striking contrast, high L2-proficiency bilinguals were faster after than before the intervention, in L1 and L2 comprehension (although in the exercise group the effect did not reach significance). According to this speculative analysis, high-proficient bilinguals would benefit from an overall practice effect, whereas slowing in low-proficient bilinguals could indicate a "fatigue" effect, or an age-related general slowing.

However, we cannot draw firm conclusions. First, L2 proficiency was not a manipulated variable, so our design did not address the question of L2 proficiency effects, and the analyses we report for low- and high-proficient groups of speakers are exploratory. Related to this, our categorization of bilinguals according to their L2 proficiency using a median split created two proficiency groups with unbalanced numbers of participants in the control and exercise groups. In fact, the high proficiency group had more exercisers (24) than controls (12), and the reverse was true for the low proficiency group (14 exercisers and 23 controls). Accordingly, there was a trend for higher proficiency in the exercise compared to the control group (means 8130.59 and 9040.26, respectively, $t = 1.80$, $p = 0.08$). Finally, it is not possible to perform mediation analysis in this even more complex design that includes what would be another moderator of the effects.

Our findings thus raise the hypothesis that L2 proficiency could play an important role in how physical exercise may affect bilingual language processing, but further studies are needed to examine whether bilingual language processing is in fact unaffected (or even negatively affected) by exercise training and its consequent improvements in CRF, and if speakers with distinct proficiency levels would be differently affected by exercise training or the passage of time. Moreover, it would be important to consider the recruitment of other cognitive functions when bilingual older adults comprehend language. Bilinguals make use of inhibitory control to avoid language interference (e.g., Green, 1998; Pivneva et al., 2012), and working memory is particularly important in language comprehension by older adults (e.g., Waters & Caplan, 2001; see also Abrams & Farrell, 2011). The effects that we reported were observed above specific linguistic domains of processing (i.e., syntax and semantics), and could thus reflect the involvement of "broader" aspects of language processing involving other functions which might be differently impacted by the exercise intervention in mono- and bilingual speakers.

Bilingualism and language dominance effects in sentence comprehension

Our manipulation of sentence context interacted with bilingualism and language effects, and this is relevant for research on bilingualism. When comparing monolinguals' to bilinguals' L1 processing (i.e., bilingualism effects), we reproduced the results from Fernandes et al. (2024), with a larger sample of older adults. Bilinguals did not employ syntactic information as effectively as monolinguals to guide comprehension to detect target words, whereas no bilingualism difference was found in the use of semantic information. In contrast, when comparing L1 to L2 processing in bilinguals (i.e., language dominance effects), we found no differences in the use of syntactic information, but the advantage of a high- compared to a low-constraint context (i.e., semantic processing) was bigger in L1 than L2.

Our results contribute to a better distinction between bilingualism and language dominance effects, which seem at times confounded in the bilingualism literature. Gollan et al. (2011;, Experiment 2) is one of the few studies that directly assessed bilingualism effects (though see Shook et al., 2015, for seemingly contrasting evidence[9]). They found no differences between monolingual and bilingual L1 comprehension of low- and high-constraint sentences, suggesting a comparable processing of semantic information. Our results also suggest identical semantic processing in comprehension by mono- and bilinguals. However, Gollan et al. (2011) compared only low- to high-constraint sentences and thus did not directly assess syntactic processing. We had an additional condition with randomly ordered word lists, which was contrasted to the low-constraint sentences, and found that bilinguals are at a disadvantage relative to monolinguals in L1 syntactic processing. This is an important result, as there is no previous direct evidence for bilingualism effects in syntactic processing, despite a general proposal of increased difficulties with syntax in bilinguals compared to monolinguals. This proposal underlies some theories of bilingualism but is based mostly on evidence from L2 processing (see Fernandes et al., 2024, for further discussion on this issue).

Concerning dominance effects, we reported a larger use of semantic constraints in L1 than in L2 comprehension. Our results are consistent with physiological evidence for a stronger reliance on semantics in L1 than L2 comprehension (Martin et al., 2013, see also Ito et al., 2017), but it should be noted that other studies failed to find differences between L1 and L2 semantic processing (Gollan et al., 2011, comparison between Spanish-English bilinguals and Dutch-English bilinguals[10]). As mentioned before, it is not clear whether semantic processing is prone to dominance effects, and divergence on results may reflect the use of different tasks and experimental manipulations.

In contrast, we failed to find differences between syntactic processing in L1 and L2. This seems inconsistent with theories for L2 processing that assume that grammar is more effortful in L2 than L1 and/or that L2 processing relies more on lexical-semantics than on syntax (e.g., Clahsen & Felser, 2006; Cunnings, 2017). We note, however, that many studies supporting those theories focused on the processing of complex syntactic structures and/or structures that maximize the influence of working memory, such as garden-path sentences (Pozzan & Trueswell, 2016) or nonlocal dependencies (i.e., sentences where an element (e.g., "Which book") must be kept in memory until upcoming information allows integration in the sentence, like "Which book did Mary ask her students to read?"; e.g., Felser and Roberts, 2007). Our results suggest that, while comprehending sentences with "normal" complexity, bilinguals are able to process syntax in L2 as efficiently as in L1, but they cannot use semantic information in L2 as much as in L1.

Overall, these findings highlight the importance of disentangling bilingualism and language dominance effects, and call for further research on how they can manifest during comprehension of sentences with varying complexity.

Limitations and future directions

It is important to note that exercise training induces physiological adaptations that we did not assess. Exercise training is known to improve physical function in respect not only to physical endurance but also resistance, reflected, for example, in increased muscle size and improved strength (Hughes et al., 2018). HIIT interventions lead to broader adaptations on the cardiovascular, respiratory, and metabolic systems, which work to improve oxygen consumption and thus CRF, making this a complex, integrative measure for physiological health (Voss & Jain, 2022). We measured $VO2_{peak}$ as an indicator of CRF, but our mediation analysis showed that this metric only explained 20% of variance in monolinguals' language comprehension. Despite $VO2_{peak}$ being a standard measure of CRF, there are other exercise-induced physiological adaptations thought to be relevant when assessing changes in aerobic fitness, especially in older adults, such as lactate threshold (LT; Fosstveit et al., 2024), which has been shown to be associated with improved cognitive performance (Jacob et al., 2023). Therefore, it is important that future studies on the effects of exercise training on language functioning consider other metrics of CRF and, potentially, other physiological measures that are known to change with exercise interventions.

Here, we employed a randomized controlled trial (RCT) with pre-post comparisons to establish causal links between exercise and language and moreover used mediation analysis to draw inferences about the mechanisms underlying those links. Relative to the traditional approach by Baron and Kenny (1986),[11] recent approaches such as the one we used are advantageous in that they allow for estimation of the amount and significance of any mediation effects. The aim of mediation analysis is to measure how the effect of a single variable (treatment) is explained by the effect of another single variable (mediator), but in RCTs with pre-post comparisons the treatment consists of the crossing of two variables, intervention group and time of intervention. We used the "mediate" function (Tingley et al., 2014), which does not allow to specify more than one treatment variable. Therefore, our treatment variable was a "condition" with four categorical levels crossing intervention group and time of intervention (Control_pre, Control_pos, Exercise_pre, Exercise_pos) with a reference level (Exercise_pre) against which the other three were contrasted. This means that, in our models, there was a direct comparison of the exercise group at pre- and post-testing, but not a direct comparison of the control group at pre- and post-testing. Moreover, the mediated effects had to be assessed for each bilingualism group separately, and for each contrast between two conditions. Whereas we do not think that this is problematic, we believe that, for the purposes of mediation analysis, future studies should try to avoid complicated designs.

Finally, it is important to consider the potential confound of language in our study and in similar designs. We discussed above our findings for bilingualism and language dominance effects on comprehension of spoken sentences that manipulated the context constraint. In our study, we had a within-subjects language dominance contrast (i.e., comparing L1 Norwegian to L2 English within the same group of participants) and a between-languages bilingualism contrast (i.e., comparing monolinguals' English to bilinguals' Norwegian). In both cases, then, two different languages are being compared, which is a potential confound, despite our efforts to match the stimuli in English and Norwegian as much as possible (see Supplemental online material 1). Therefore, replication of these findings holding language constant and testing different groups of speakers would be important to avoid such confounding (although this introduces instead variability at the participant level, which is in our design avoided for language dominance effects).

Conclusion

We provided evidence supporting the hypothesis that exercise training is beneficial for language comprehension by healthy older adults. We showed that speeded comprehension in monolinguals was explained by exercise-induced increased $VO2_{peak}$. However, this was not observed in bilinguals, contradicting our prediction that the exercise-related advantages might be bigger in the more costly bilingual language processing. Exploratory analyses pointed to differences in comprehension between low- and high- L2 proficient bilinguals, but future research needs to address this question in detail. Finally, we provided evidence for bilingualism effects whereby bilinguals are at a disadvantage, relative to monolinguals, in syntactic processing, and for language dominance effects, whereby in L1 processing there is more use of semantic information to guide comprehension than in L2.

Notes

1. Note that the detrimental effects of aging in language comprehension seem to be observed equally in mono- and bilinguals, that is, bilingualism and aging independently affect language. The current study focuses on mono- and bilingual older adults, who have both been shown to have poorer comprehension than their younger counterparts (Fernandes et al., 2024).
2. However, if participants had a partner (whom they lived with) also taking part in the study, they would be assigned to the same group.
3. Two end criteria used to assess the validity of VO2peak testing were RER(VCO2/VO2) and blood lactate concentration (mmol/L). For the present sample, RER was 1.09 ± 0.06, which is above 1, the value recommended for this end criteria for older adults (Edvardsen et al., 2014). Blood lactate concentration (mmol/L) was 7.48 ± 1.88. For this end criteria the figures recommended are ≥ 3.5 for females and 4.00 for males (Edvardsen et al., 2014). Measured blood lactate concentration in the female exercisers was 7.28 ± 1.8 and in the male exercisers was 7.70 ± 1.96.
4. Norwegian has two official written languages – Bokmål and Nynorsk. Bokmål was used in this study since the vast majority of writing is done in bokmål (85–90%): Language Council of Norway (https://www.sprakradet.no/Spraka-vare/Norsk/fakta-om-norsk/.
5. Note that, for the purpose of the mediation analysis, the intervention was coded as a 4-level variable and its contrasts never compare directly performance pre- and post-intervention for the control group, but both these conditions against pre-intervention in the exercise group. Therefore, we should retain from the group analysis that, for bilinguals, there was a slowing from pre- to post-testing, across intervention groups.
6. The variable “cond” is fed to the “treat” argument of the “mediate” function, and the “control. value” and “treat.value” options are used to calculate the specific contrasts of interest. The direct and mediated (through VO2peak) effects are calculated subsequently for each level of the moderator (monolinguals and bilinguals), which are specified in the “covariates” argument (Tingley et al., 2014).
7. Note that the measurements of maximum oxygen were taken at the times of pre- and post-testing, and at each point in time language processing data were collected for both L1 and

L2. Therefore, VO2peak data for bilinguals is the same data as analyzed in the comparison monolinguals-bilinguals, and not affected by the variable Language (as shown in the mediator model).

8. However, it is interesting to note that in L2 the slowing in the control group was weaker and of smaller magnitude, which resembles the finding for slowing restricted to the exercise group in L2 comprehension.
9. Shook et al. (2015) found a trend for English-German bilinguals to differ from English monolinguals in fixations to pictures of target words embedded in (spoken) high-constraint sentences. This was interpreted as a smaller effect of a predictable sentence context by bilinguals than monolinguals. However, on one hand, this was only a trend, and, on the other hand, this task included the depiction of (L2) competitor words, which may have contributed to these effects by increasing the proportion of fixations to a competitor and, consequently, by decreasing proportion of fixations to the target, in bilinguals.
10. We note that a direct comparison with Gollan et al. (2011) is difficult, first and foremost because their study more directly tapped into frequency effects, which were assessed in the different constraint and speakers' groups conditions. Also, the focus of their follow-up analyses was on L1-L2 differences in low-constraint and in high-constraint, separately, whereas we focused on the low vs. high-constraint contrast as a measure of semantic processing in L1 and L2 processing separately. The same argument applies relative to a comparison with Shook et al.'s (2015) results. Moreover, we note that Shook et al. (2015) compared L2 processing to monolinguals' L1 processing, which is not a direct assessment of language dominance effects.
11. The standard procedure for mediation analysis is to fit a set of linear regressions, from which the role of the mediator is inferred: (i) regressing the mediator on the IV, (ii) regressing the dependent variable on the IV, and (iii) regressing the dependent variable on both the IV and on the mediator (Baron & Kenny, 1986). If the effects of the IV can be explained by mediation, they should disappear when the mediator is added to the model (i.e., in regression (iii) compared to regression (ii)).

Acknowledgments

We would like to thank the research assistants who helped to conduct the data collection: Ali Kalib Ali, Renate Gjetnes, Consuelo Vidal Gran, Amelie Grandjean, Nicolas Hayston Lecaros, Karethe Nilsen, Helene Øye, Øystein Søren Østerhus, Rupali Limachya, Kelsey E. Joyce, and Ellinor Skjerli. We also thank Jan Zandhuis for technical support.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Research Council of Norway under [Grant number 286961].

ORCID

Eunice G. Fernandes http://orcid.org/0000-0002-1448-3256
Katrien Segaert http://orcid.org/0000-0002-3002-5837

Data availability statement

The data that support the findings of this study are openly available in OSF at https://osf.io/d7aw2/.

References

Abrams, L., & Farrell, M. T. (2011). Language processing in normal aging. In J. Guendouzi, F. Loncke, & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes: Perspectives in communication disorders* (pp. 49–73). Psychology Press. https://doi.org/10.4324/9780203848005ch3

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/jjml200712005

Balke, B., & Ware, R. W. (1959). An experimental study of physical fitness of air force personnel. *United States Armed Forces Medical Journal*, *10*(6), 675–688.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459–464. https://doi.org/10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effects of bilingualism on cognitive control in the Simon task: Evidence from MEG. *Neuroimage*, *24*(1), 40–49. https://doi.org/10.1016/j.neuroimage.2004.09.044

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, *2*(2), 92–98. https://doi.org/10.2340/1650197719702239298

Bouaziz, W., Malgoyre, A., Schmitt, E., Lang, P. O., Vogel, T., & Kanagaratnam, L. (2020). Effect of high-intensity interval training and continuous endurance training on peak oxygen uptake among seniors aged 65 or older: A meta-analysis of randomized controlled trials. *International Journal of Clinical Practice*, *74*(6). https://doi.org/10.1111/ijcp.13490

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–252. https://www.jstor.org/stable/2984418

Brown, B. M., Frost, N., Rainey-Smith, S. R., Doecke, J., Markovic, S., Gordon, N., Weinborn, M., Sohrabi, H. R., Laws, S. M., Martins, R. N., Erickson, K. I., & Peiffer, J. J. (2021). High-intensity exercise and cognitive function in cognitively normal older adults: A pilot randomised clinical trial. *Alzheimer's Research & Therapy*, *13*(1), 33. https://doi.org/10.1186/s13195-021-00774-y

Burzynska, A. Z., Chaddock-Heyman, L., Voss, M. W., Wong, C. N., Gothe, N. P., Olson, E. A., Knecht, A., Lewis, A., Monti, J. M., Cooke, G. E., Wojcicki, T. R., Fanning, J., Chung, H. D., Awick, E., McAuley, E., Kramer, A. F., & Margulies, D. (2014). Physical activity and cardiorespiratory fitness are beneficial for white matter in low-fit older adults. *PLOS ONE*, *9*(9), e107413. https://doi.org/10.1371/journal.pone.0107413

Campbell, K. L., Samu, D., Davis, S. W., Geerligs, L., Mustafa, A., & Tyler, L. K. (2016). Robust resilience of the frontotemporal syntax system to aging. *Journal of Neuroscience*, *36*(19), 5214–5227. https://doi.org/10.1523/JNEUROSCI.4561-15.2016

Caplan, D., De de, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging*, *26*(2), 439–450. https://doi.org/10.1037/a0021837

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal cognitive assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry*, *33*(2), 379–388. https://doi.org/10.1002/gps4756

Cashin, A. G., & Lee, H. (2021). An introduction to mediation analyses of randomized controlled trials. *Journal of Clinical Epidemiology*, *133*, 161–164. https://doi.org/10.1016/j.jclinepi.2021.02.014

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports*, *100*(2), 126–131.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Colcombe, S. J., Erickson, K. I., Scalf, P. E., Kim, J. S., Prakash, R., McAuley, E., Elavsky, S., Marquez, D. X., Hu, L., & Kramer, A. F. (2006). Aerobic exercise training increases brain volume in aging humans. *Journals of Gerontology: Series A*, *61*(11), 1166–1170. https://doi.org/10.1093/gerona/61.11.1166

Colcombe, S. J., Kramer, A. F., Erickson, K. I., Scalf, P., McAuley, E., Cohen, N. J., Webb, A., Jerome, G. J., Marquez, D. X., & Elavsky, S. (2004). Cardiovascular fitness, cortical plasticity, and aging. *Proceedings of the National Academy of Sciences*, *101*(9), 3316–3321. https://doi.org/10.1073/pnas.0400266101

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 659–678. https://doi.org/10.1017/S1366728916000675

Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders*, *20*(3), 245–257. https://doi.org/10.1016/0021-9924(87)90022-0

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*(4), 850–855. https://doi.org/10.3758/PBR.15.4.850

Edvardsen, E., Hem, E., & Anderssen, S. A. (2014). End criteria for reaching naximal oxygen uptake must be strict and adjusted to sex and age: A cross-sectional study. *PLOS ONE*, *9*(1), e85276. https://doi.org/10.1371/journal.pone.0085276

Etnier, J. L., Nowell, P. M., Landers, D. M., & Sibley, B. A. (2006). A meta-regression to examine the relationship between aerobic fitness and cognitive performance. *Brain Research Reviews*, *52*(1), 119–130. https://doi.org/10.1016/j.brainresrev.2006.01.002

Falck, R. S., Davis, J. C., Best, J. R., Crockett, R. A., & Liu-Ambrose, T. (2019). Impact of exercise training on physical and cognitive function among older adults: A systematic review and meta-analysis. *Neurobiology of Aging*, *79*, 119–130. https://doi.org/10.1016/j.neurobiolaging.2019.03.007

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1111/1469-8986.3920133

Federmeier, K. D., van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging*, *18*(4), 858–872. https://doi.org/10.1037/0882-7974.18.4.858

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, *23*(1), 9–36. https://doi.org/10.1177/0267658307071600

Fernandes, E. G., Segaert, K., Rahman, F., Wetterlin, A., & Wheeldon, L. (2024). Bilingualism and ageing independently impact on language processing: Evidence from comprehension and production. *Bilingualism: Language and Cognition*, 1–15. https://doi.org/10.1017/S1366728924000245

Feron, J., Segaert, K., Rahman, R., Fosstveit, S., Joyce, K. E., Gilani, A., Lohne-Seiler, H., Berntsen, S., Mullinger, K. J., & Lucas, S. J. E. (2024). Determinants of cerebral blood flow and arterial transit time in healthy older adults. *Aging (Albany NY)*, *16*(18), 12473–12497. https://doi.org/10.18632/aging.206112

Fosstveit, S., Berntsen, S., Feron, J., Joyce, K., Ivarsson, A., Segaert, K., Lucas, S., & Lohne-Seiler, H. (2024). HIIT at home: Enhancing cardiorespiratory fitness in older adults—A randomized controlled trial. *Scandinavian Journal of Medicine & Science in Sports*, *34*(7), e14694. https://doi.org/10.1111/sms.14694

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology. General*, *140*(2), 186–209. https://doi.org/10.1037/a0022256

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

Haeger, A., Costa, A. S., Schulz, J. B., & Reetz, K. (2019). Cerebral changes improved by physical activity during cognitive decline: A systematic review on MRI studies. *NeuroImage Clinical*, *23*, 101933. https://doi.org/10.1016/j.nicl.2019.101933

Howell, D. (2006). *Statistical methods for psychology* (6th ed.). Wadsworth.

Hughes, D. C., Ellefsen, S., & Baar, K. (2018). Adaptations to endurance and strength training. *Cold Spring Harbor Perspectives in Medicine*, *8*(6), a029769. https://doi.org/10.1101/cshperspect.a029769

Intzandt, B., Vrinceanu, T., Huck, J., Vincent, T., Montero-Odasso, M., Gauthier, C. J., & Bherer, L. (2021). Comparing the effect of cognitive vs. exercise training on brain MRI outcomes in healthy older adults: A systematic review. *Neuroscience & Biobehavioral Reviews*, *128*, 511–533. https://doi.org/10.1016/j.neubiorev.2021.07.003

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, *32*(8), 954–965. https://doi.org/10.1080/23273798.2016.1242761

Jacob, N., So, I., Sharma, B., Marzolini, S., Tartaglia, M. C., Oh, P., & Green, R. (2023). Effects of high-intensity interval training protocols on blood lactate levels and cognition in healthy adults: Systematic review and meta-regression. *Sports Medicine*, *53*(5), 977–991. https://doi.org/10.1007/s40279-023-01815-2

Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition*, *18*(5), 655–681. https://doi.org/10.1080/13506280902986058

Kramer, A. F., & Erickson, K. I. (2007). Capitalizing on cortical plasticity: Influence of physical activity on cognition and brain function. *Trends in Cognitive Sciences*, *11*(8), 342–348. https://doi.org/10.1016/j.tics.2007.06.009

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

MacKinnon, D. P., Cheong, J., & Pirlott, A. G. (2012). Statistical mediation analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology*. Research designs: Quantitative, qualitative, neuropsychological, and biological (Vol. 2, pp. 313–331). American Psychological Association. https://doi.org/10.1037/13620-018

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, & Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marslen-Wilson, W. D., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature*, *257*, 784–786. https://doi.org/10.1038/257784a0

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/jjml201308001

Maylor, E. A. (1990). Age, blocking and the tip of the tongue state. *British Journal of Psychology*, *81*, 123–134. https://doi.org/10.1111/j.2044-8295.1990.tb02350.x

Menenti, L., Segaert, K., & Hagoort, P. (2012). The neuronal infrastructure of speaking. *Brain and Language*, *122*(2), 71–80. https://doi.org/10.1016/j.bandl.2012.04.012

Montoya, A. K., Hayes, A. F., & Gomez Juarez, N. (2023, September 21). Mediation analysis in the two-group pretest-posttest design: A new approach based on treatment as a moderator of Time effects. *PsyArxiv Preprints*. https://doi.org/10.31234/osf.io/rfvs7

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j1532-5415200553221x

Northey, J. M., Cherbuin, N., Pumpa, K. L., Smee, D. J., & Rattray, B. (2018). Exercise interventions for cognitive function in adults older than 50: A systematic review with meta-analysis. *British Journal of Sports Medicine*, *52*(3), 154–160. https://doi.org/10.1136/bjsports-2016-096587

Ossher, L., Flegal, K. E., & Lustig, C. (2013). Everyday memory errors in older adults. *Aging, Neuropsychology & Cognition*, *20*(2), 220–242. https://doi.org/10.1080/13825585.2012.690365

Peelle, J. E. (2018). Language and aging. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The oxford handbook of neurolinguistics* (pp. 295–316). Oxford University Press.

Perani, D. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain A Journal of Neurology*, *121*(10), 1841–1852. https://doi.org/10.1093/brain/121.10.1841

Perani, D., Dehaene, S. G., Cohen, F., Cappa, L., F, S., Dupoux, E., Fazio, F., & Mehler, J. (1996). Brain processing of native and foreign languages. *NeuroReport*, *7*(15), 2439–2444. https://doi.org/10.1097/00001756-199611040-00007

Pivneva, I., Palmer, C., & Titone, D. (2012). Inhibitory control and L2 proficiency modulate bilingual language production: Evidence from spontaneous monologue and dialogue speech. *Frontiers in Psychology*, *3*, 3. https://doi.org/10.3389/fpsyg.2012.00057

Poulisse, C., Wheeldon, L. R., & Segaert, K. (2019). Evidence against preserved syntactic comprehension in healthy aging. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *45*(12), 2290–2308. https://doi.org/10.1037/xlm0000707

Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: A visual word study a visual word study. *Bilingualism: Language and Cognition*, *19*(3), 636–643. https://doi.org/10.1017/s1366728915000838

Rahman, F. T., A, K., Feron, J., Mullinger, K., Joyce, K., Gilani, A., Fernandes, E. G., Wetterlin, A., Wheeldon, L., Lucas, S. J. E., & Segaert, K. (2024 2023). Lifestyle and brain health determinants of word-finding failures in healthy ageing. *bioRxiv*. https://doi.org/10.1101/2023.12.08.570799

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Rijnhart, J. J. M., Twisk, J. W. R., Valente, M. J., & Heymans, M. W. (2022). Time lags and time interactions in mixed effects models impacted longitudinal mediation effect estimates. *Journal of Clinical Epidemiology*, *151*, 143–150. https://doi.org/10.1016/j.jclinepi.2022.07.004

Rosano, C., Venkatraman, V. K., Guralnik, J., Newman, A. B., Glynn, N. W., Launer, L., Taylor, C. A., Williamson, J., Studenski, S., Pahor, M., & Aizenstein, H. (2010). Psychomotor speed and functional brain MRI 2 years after completing a physical activity treatment. *Journals of Gerontology: Series A*, *65A*(6), 639–647. https://doi.org/10.1093/gerona/glq038

Saur, D., Baumgaertner, A., Moehring, A., Büchel, C. B., Musso, M., Rose, M., M, Meisel, J. M., & Musso, M. (2009). Word order processing in the bilingual brain. *Neuropsychologia*, *47*(1), 158–168. https://doi.org/10.1016/j.neuropsychologia.2008.08.007

Segaert, K., Lucas, S. J. E., Burley, C. V., Segaert, P., Milner, A. E., Ryan, M., & Wheeldon, L. (2018). Higher physical fitness levels are associated with less language decline in healthy ageing. *Scientific Reports*, *8*(1), 6715. https://doi.org/10.1038/s41598-018-24972-1

Shook, A., Goldrick, M., Engstler, C., & Marian, V. (2015). Bilinguals show weaker lexical access during spoken sentence comprehension. *Journal of Psycholinguistic Research*, *44*(6), 789–802. https://doi.org/10.1007/s10936-014-9322-6

Smith, P. J., Blumenthal, J. A., Hoffman, B. M., Cooper, H., Strauman, T. A., Welsh-Bohmer, K., Browndyke, J. N., & Sherwood, A. (2010). Aerobic exercise and neurocognitive performance: A meta-analytic review of randomized controlled trials. *Psychosomatic Medicine*, *7*(3), 239–252. https://doi.org/10.1097/psy.0b013e3181d14633

Sulpizio, S., Del Maschio, N., Del Mauro, G., Fedeli, D., & Abutalebi, J. (2020). Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *Neuroimage*, *205*, 116306. https://doi.org/10.1016/j.neuroimage.2019.116306

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2013). *Mediation: R package for causal mediation analysis. R package version 4.4.2*, http://CRAN.R-project.org/package=mediation

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: The modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex*, *20*(2), 352–364. https://doi.org/10.1093/cercor/bhp105

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer.

Vidoni, E. D., Johnson, D. K., Morris, J. K., Van Sciver, A., Greer, C. S., Billinger, S. A., Donnelly, J. E., Burns, J. M., & Fukumoto, Y. (2015). Dose-response of aerobic exercise on cognition: A community-based, pilot randomized controlled trial. *PLOS ONE*, *10*(7), e0131647. https://doi.org/10.1371/journal.pone.0131647

Vingerhoets, G. (2003). Multilingualism: An fMRI study. *Neuroimage*, *20*(4), 2181–2196. https://doi.org/10.1016/j.neuroimage.2003.07.029

Voss, M. W., Heo, S., Prakash, R. S., Erickson, K. I., Alves, H., Chaddock, L., Szabo, A. N., Mailey, E. L., Wójcicki, T. R., White, S. M., Gothe, N., McAuley, E., Sutton, B. P., & Kramer, A. F. (2013). The influence of aerobic fitness on cerebral white matter integrity and cognitive function in older adults: Results of a one-year exercise intervention. *Human Brain Mapping*, *34*(11), 2972–2985. https://doi.org/10.1002/hbm.22119

Voss, M. W., & Jain, S. (2022). Getting fit to counteract cognitive aging: Evidence and future directions. *Physiology*, *37*(4), 197–206. https://doi.org/10.1152/physiol.00038.2021

Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging*, *16*(1), 128–144. https://doi.org/10.1037/0882-7974.16.1.128