Title: Fit to speak - Physical fitness is associated with reduced language decline in healthy ageing

Authors:

K. Segaert[1,2*], S.J.E. Lucas[3,2], C.V. Burley[3], P. Segaert[4], A. E. Milner[1], M. Ryan[5], L. Wheeldon[6,1]

Affiliations:

[1] School of Psychology, University of Birmingham, Birmingham, UK

[2] Centre for Human Brain Health, University of Birmingham, Birmingham, UK

[3] School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, UK

[4] Department of Mathematics, University of Leuven, Leuven, Belgium

[5] Cardiovascular Division, King's College London, London, UK

[6] Department of Foreign Languages and Translation, University of Agder, Kristiansand, Norway

* Address correspondence to:

Dr. Katrien Segaert

School of Psychology

University of Birmingham

Edgbaston

Birmingham

B15 2TT

UK

k.segaert@bham.ac.uk

Abstract

Healthy ageing is associated with decline in cognitive abilities such as language. Aerobic fitness has been shown to ameliorate decline in some cognitive domains, but the potential benefits for language have not been examined. We investigated the relationship between aerobic fitness and tip-of-the-tongue states. These are among the most frequent cognitive failures in healthy older adults and occur when a speaker knows a word but is unable to produce it. We found that healthy older adults indeed experience more tip-of-the-tongue states, and that when they do they have access to less information about the word's sound structure, compared to young controls. Importantly, higher aerobic fitness levels decrease the probability of experiencing tip-of-the-tongue states in healthy older adults, over and above the effect of age. This is the first demonstration of a link between aerobic fitness and language functioning in healthy older adults.

Keywords: aerobic fitness, brain health, ageing, language, tip-of-the-tongue

Introduction

If the existing demographic trends continue, then in countries with high life expectancies such as the UK, Canada, the US and Japan, most children born after the year 2000 will live to become 100 years old[1]. In part motivated by the economic, healthcare and social challenges associated with this shift, there is an ever-growing interest in uncovering the antecedents of healthy ageing. In this paper we focus on cognitive changes, particularly changes in language abilities, in the healthy ageing population. Temporary cognitive lapses, such as not having a word come to mind when speaking, occur more frequently as we grow older. An interesting question therefore is whether lifestyle factors are related to the occurrence of such 'senior moments'.

Maintaining good language skills is important for older adults. The experience of healthy ageing is only loosely related to one's numerical age. The way people move, feel, think, interact with and talk to others all co-determine the ageing experience. Focus groups have highlighted that maintaining social relations and independence are particularly instrumental for a positive experience of ageing[2], and good language abilities are crucial for achieving this. It is therefore also not surprising that when asked about age-related cognitive failures, older adults report that word finding difficulties are particularly irritating and embarrassing[3].

Word finding difficulties often surface as tip-of-the-tongue experiences. People in a tip-of-the-tongue state have a strong conviction that they know a word, but are unable to produce it. The frequency of tip-of-the-tongue states increases with age[4] and indeed tip-of-the-tongue states are documented to be among the most frequent cognitive failures in healthy older adults[5]. Older adults worry that tip-of-the-tongue states indicate serious memory problems[3]. However, this is a misconception: tip-of-the-tongue states are not associated with episodic memory loss[6]. In fact, older adults usually have a much larger vocabulary than younger adults[7]. Instead, focused experimental research has demonstrated that tip-of-the-tongue states are indicative of deficits in accessing phonology (i.e. sound form representations). Spoken word production is a two-stage process involving the retrieval of word meaning, followed by the associated phonology[8]. Tip-of-the-tongue states indicate a disruption in the process of transmission between meaning and phonology[4,9,10]. This process is essential for successful and fluent language production, and its disruption has very noticeable negative consequences for elderly speakers.

In the present research we investigated whether older adults' aerobic fitness levels are related to the incidence of age-related language failures such as tip-of-the-tongue states. Regular physical exercise is a lifestyle intervention approach that has received a considerable amount of attention in previous studies. As an intervention strategy, regular exercise is accessible, safe and effective[11]. Interventions as short as 6 weeks can result in a measurable increase in aerobic fitness[12]. Moreover, at least some modes of regular physical exercise are easily accessible, such as walking. Even when the primary mode of exercise that is taken up is walking, research has demonstrated short- and long-term cognitive benefits[13]. To date, no studies have investigated whether there is a relationship between aerobic fitness and language functioning. This stands in stark contrast to the amount of evidence of aerobic fitness benefits

for other cognitive domains (e.g., cognitive control, executive functioning, visuo-spatial memory, learning and processing speed)[14,15].

Regular physical exercise and the resultant higher aerobic fitness is associated with reducing age-related decline in brain perfusion[16] and structural integrity[17]. Among others, structural integrity in frontal and temporal regions of the brain has been related to aerobic fitness[18,19]. This leads us to hypothesize that cognitive benefits of aerobic fitness may extend to the domain of language processing: language production is predominantly associated with functional activation in frontal and temporal regions in the brain[20]; word finding difficulties in particular are associated with functional activation[21] and structural atrophy[22] of the left insula. All this suggests that aerobic fitness may be associated with age-related decline in word finding abilities; however to-date, no study has generated empirical evidence to support this hypothesis.

In the present study, we investigated the relationship between aerobic fitness and word finding abilities in a cross-sectional sample of healthy older adults. Word finding abilities were measured in a tip-of-the-tongue-eliciting experiment, in which the participants read definitions of words and were asked to produce the word. Aerobic fitness was quantified using a physiological measure of oxygen uptake from a graded exercise test. Using this gold-standard objective measurement of aerobic fitness, rather than a less reliable self-report physical activity measurement[23], we were able to accurately assess individual aerobic fitness levels. We hypothesized that there would be a positive relation between aerobic fitness levels and word finding abilities in our sample of healthy older adults, independent of age and vocabulary size. A demonstration of such a link could have far-reaching implications: the promise of potential ameliorating effects of regular physical exercise on an important and complex cognitive ability - language production - would add a crucial piece of knowledge to this growing and timely area of research.

Results

Older adults experience more tip-of-the-tongue occurrences and have less access to phonological information, compared to young participants

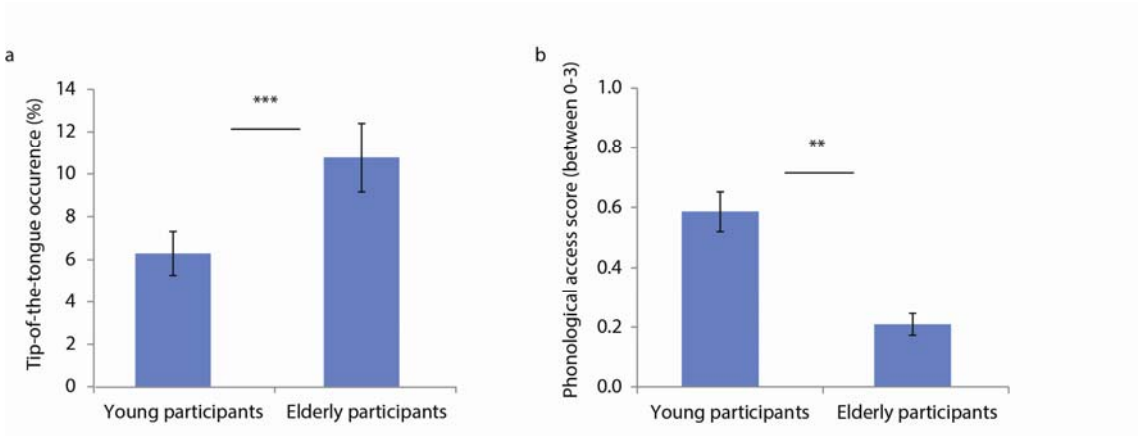

Figure 1. Age-related decline in word finding abilities. (a) There was a significant increase in the percentage of tip-of-the-tongue states experienced by the older adults compared to the young participants. (b) When experiencing tip-of-the-tongue states, older adults had access to significantly less correct phonological information than young participants.

Figure 1 summarizes the age-related decline in healthy older adults compared to young controls with regard to word finding abilities.

In line with our predictions, we found that older adults experienced more tip-of-the-tongue occurrences than young participants ($p<0.001$) (Figure 1a).

Both young and older adults experienced fewer tip-of-the-tongue occurrences the shorter the target words (as measured by the number of phonemes) ($p<0.001$) and the larger their vocabulary size ($p<0.001$).

Next we looked at the age-related decline in the access of correct phonological information about the words. Older adults had less access to correct phonological information than young participants ($p<0.002$) (Figure 1b).

Also in line with previous observations, older adults had a significantly larger vocabulary than young participants ($t(53)=5.01$, $p<0.001$).

Table 1 summarises the fitted mixed effects logistic regression model we used for the tip-of-the-tongue occurrences and the mixed effects linear regression model used on phonological access scores in older adults against our control group of younger participants.

Table 1. Summary of the mixed effects regression models predicting tip-of-the-tongue occurrence and phonological access in young versus older adults

A. Mixed effects logistic regression model predicting tip-of-the-tongue occurrence

	Coefficient	*SE*	Walds *z*	*p* value	
Intercept	-3.51	0.24	-14.52	<0.001	***
Age group	-0.75	0.14	-5.24	<0.001	***
No. phonemes of target word	0.30	0.08	3.60	<0.001	***
Vocabulary size	-0.06	0.01	-4.30	<0.001	***

Note: N = 3300, AIC = 1699.5, log-likelihood = -839.8
This model includes a random intercept for items and participants, a random slope for Vocabulary Size for items and a random slope for the Number of Phonemes of the Target Word for participants. Multicollinearity was low (all VIF < 1.4)

B. Mixed effects linear regression model predicting phonological access scores

	Coefficient	*SE*	*df*	t value	*p* value	
Intercept	0.51	0.08	47.31	6.38	<0.001	***
Age group	0.24	0.07	38.01	3.32	<0.002	**

Note: N = 272, AIC = 568.58.
This model includes a random intercept for items and participants
*** < .001 ** < .01 * < .05

Aerobic fitness level is positively related to less frequent tip-of-the-tongue occurrence in healthy older adults

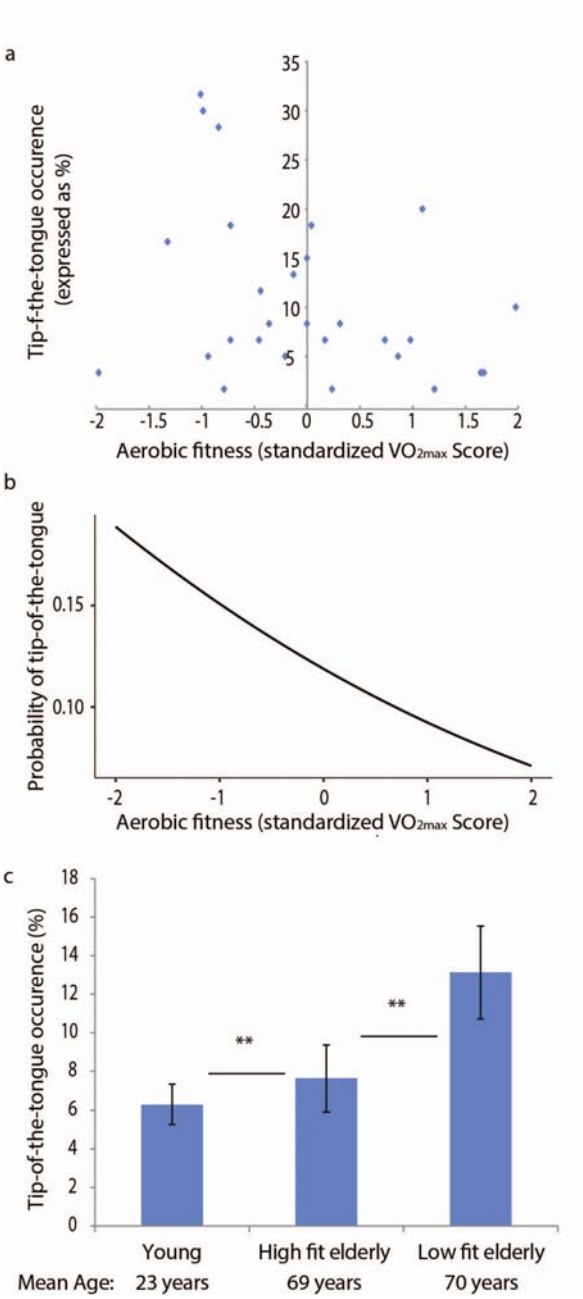

Figure 2. Aerobic fitness ameliorates age-related decline in word finding abilities. (a) Depicted for older adults is the tip-of-the-tongue occurrence as a function of aerobic fitness scores. (b) There was a significant influence of the standardized aerobic fitness scores on the probability of experiencing a tip-of-the-tongue state. Depicted is the mean effect across participants, keeping all other variables in the model (see table 2) constant. The higher the aerobic fitness score, the lower the probability of experiencing a tip-of-the-tongue state. (c) When the older adults were divided into groups of high- and low-fitness (high- and low-fit as determined by a median split, groups were matched for age, education level and vocabulary size), we found that low-fit older adults had a higher incidence of tip-of-the-tongue states than high-fit older adults. High-fit older adults in turn had a higher incidence than the young participants.

Next we investigated the relationship between aerobic fitness and word finding abilities in healthy older adults.

First, we tested whether aerobic fitness scores were predictive of the probability of experiencing a tip-of-the-tongue state. Our objective physiological marker of aerobic fitness ($V\square O_{2max}$) indeed significantly predicted tip-of-the-tongue occurrences ($p=0.014$). Thus, the more aerobically fit the older adults were, the less likely they were to experience a tip-of-the-

tongue state. Figure 2b depicts the influence of aerobic fitness scores on the probability of experiencing a tip-of-the-tongue state, keeping all other variables in the model constant (the figure shows the mean effect across participants).

The relationship between aerobic fitness and tip-of-the-tongue occurrence was observed over and above the effects of age and vocabulary size. Effects of age and vocabulary size were also accounted for in the model. Tip-of-the-tongue occurrences increased with participant age (p=0.008). Also, the larger the vocabulary size of the older adults, the less likely they were to have tip-of-the-tongue experiences (p=0.005). Older adults experience more tip-of-the-tongue occurrences for longer words, as measured by the number of phonemes of the target word (p=0.004). Education level did not contribute toward predicting tip-of-the-tongue states (this may be because there was relatively little variation in education level among our participants, with only a few who had not received formal education at university level).

The result of the mixed effects logistic regression model of tip-of-the-tongue occurrence in the group of older adults is summarized in table 2.

A mixed effects linear regression model on phonological access scores, revealed that aerobic fitness scores were not predictive of phonological access (p>.8).

Lastly, we performed a median split on the standardized aerobic fitness scores to create a high-fit older adults and a low-fit older adults group. Comparison of group means revealed that the two groups did not differ in age (no assumption of equal variance: $t_{(21.17)}$=-1.05,p>0.3), vocabulary size ($t_{(26)}$=.20,p>0.8) or education level ($t_{(26)}$=-.60,p>0.5). Despite the fact that high-fit and low-fit older adults are thus matched on age, education level and vocabulary size, the tip-of-the-tongue occurrence of high-fit older adults was lower than that of low-fit older adults (p<.004), although still higher than that of young adults (p<0.003) (Figure 2c).

In sum, the data show a relationship between aerobic fitness and word finding abilities in a group of healthy older adults.

Table 2. Summary of the mixed effects logistic regression model predicting tip-of-the-tongue occurrence in the older adults

	Coefficient	*SE*	Walds *z*	*p* value	
Intercept	-3.4	0.30	-11.38	<0.001	***
No. phonemes of target word	0.3	0.10	2.89	0.004	**
Vocabulary size	-0.06	0.02	-2.80	0.005	**
Age	0.06	0.02	2.68	0.008	**
Aerobic fitness ($V\square O_{2max}$ score)	-0.3	0.12	-2.47	0.014	*

Note: N = 1680, AIC = 989.8, log-likelihood = -483.9, *** < .001 ** < .01 * < .025
This model included a random intercept for items and participants, a random slope for Vocabulary Size for items and a random slope for the Number of Phonemes of the Target Word for participants. Multicollinearity was low (all VIF < 1.1).

Power of detecting an aerobic fitness effect on tip-of-the-tongue occurrence in healthy older adults

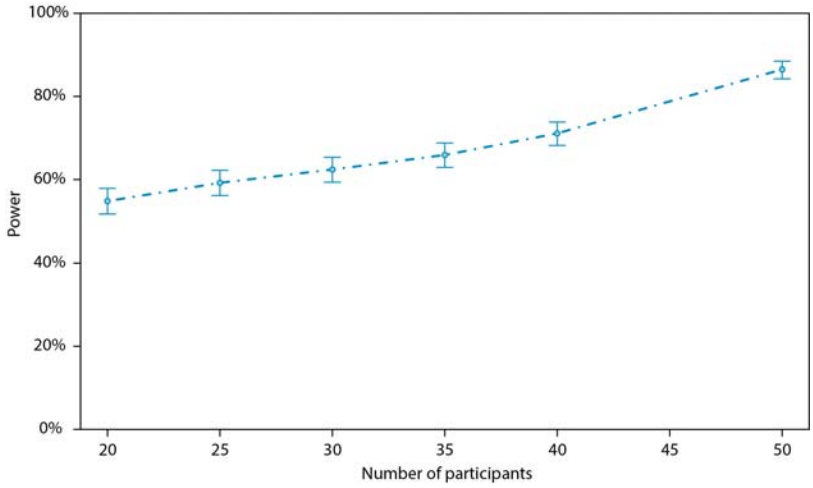

Figure 3. Power of the aerobic fitness effect on tip-of-the-tongue occurrence in function of the number of participants. Power is estimated as the chance of detecting a non-zero effect of aerobic fitness on the probability of experiencing a tip-of-the-tongue state in a simulation study with 1000 simulation runs. The figure indicates that already a small sample size leads to a high power of the study.

The chance of a Type-I error (i.e. incorrectly rejecting the null-hypothesis) for all above described effects is fixed to 2.5 %, since all effects were tested on the 2.5% level (i.e. p-values < 0.025). With the aim to provide information on Type-II error in possible future studies, we investigated the power of detecting a non-zero effect of aerobic fitness on the probability of experiencing a tip-of-the-tongue state by means of a simulation study. Under the above fitted model, we simulated new independent data sets with a given number of participants using the R-package simr[24]. The power then signals the percentage of those simulated data sets for which the hypothesis of a zero effect is rejected. Figure 3 shows the obtained power for 1000 simulation runs in function of the number of participants included in the study. For a sample of 25 and 30 participants, a respective power of 59.20% (confidence interval [56.08, 62.27]) and 62.40% (confidence interval [59.31, 65.41]) is obtained. Note that our current study uses N = 28 participants. We can therefore conclude that the effect size of the aerobic fitness effect on tip-of-the-tongue occurrence would lead to what the field of psychology would consider a medium power effect even for relatively small sample sizes. Figure 3 can be used to inform sample size estimation and power analyses in future studies.

Discussion

We demonstrate for the first time that there is a relationship between language production abilities and aerobic fitness in healthy older adults. The data reveal that healthy older adults have more tip-of-the tongue occurrences, and when they do, have less access to phonological information about the target word in comparison to the young control group. However, most

importantly, higher aerobic fitness levels are associated with better word finding abilities in older adults. Specifically, there is a relation between tip-of-the-tongue occurrence and aerobic fitness over and above the influence of age and vocabulary size. The higher the older adults' aerobic fitness level, the lower the probability of experiencing a tip-of-the-tongue state.

There is an increasing interest in lifestyle factors that could ameliorate age-related decline. A growing number of studies have demonstrated a relationship between regular physical exercise and different domains of cognition[14], but the present study is to the authors' knowledge the first to investigate the link between fitness and language processing. Language is a crucial aspect of cognition, necessary for maintaining independence, communication and social interaction in older age. Clearly, language processing is related to general cognitive functions such as processing speed, executive functions, working memory and declarative memory. However, language functioning cannot be reduced to functioning in these non-linguistics domains and the relationship between language decline and the decline in these non-linguistics cognitive functions is not yet fully understood. It is therefore noteworthy that we are able to show, for the first time, that the benefits of aerobic fitness extend to the domain of language.

Our finding that older adults experience more tip-of-the-tongue states, and have less access to correct phonological information when in a tip-of-the-tongue state is in line with previous observations[4,9,10]. Previous research on the impact of ageing on language processing has identified that older adults have reduced word finding abilities, and also have syntactic processing difficulties when speaking. Consequently, older adults choose to produce less complex syntactic structures and make more grammatical errors, e.g. [25]. One interesting implication from the findings of the current study is that future research could explore whether higher aerobic fitness levels also relate to age-related decline in this aspect of language processing.

Considerable progress has been made in elucidating the changes in brain function and structure that underlie the cognitive benefits associated with higher levels of physical activity, which ultimately manifest in increases in aerobic fitness. Firstly, greater aerobic fitness is associated with vascular benefits for the ageing brain. Effective regulation of brain blood flow, including the effectiveness of blood vessels to respond to changes in the concentration of carbon dioxide (CO_2), is vital for optimal brain function[11,26]. Cerebral blood flow and cerebrovascular reactivity to CO_2 decrease with age, but higher aerobic fitness levels are associated with less decline in these measures of brain function[16]. Moreover, vascular function has been shown to improve following an exercise training intervention aimed at increasing aerobic fitness[27]. Several studies have demonstrated a relationship between vascular brain functioning and various aspects of cognitive performance, including cognitive control[28] and speed[29]. Secondly, aerobic fitness is related to neural effects in the ageing brain. Healthy ageing leads to reductions in grey matter volume and altered integrity of white matter tracts[30], however higher aerobic fitness levels have been demonstrated to act as a protector for age-related decline in the brain's structural integrity[17]. Higher aerobic fitness is associated with greater grey matter volume in frontal and hippocampal regions[31]. Moreover exercise training in sedentary older adults is demonstrated to have widespread effects, sparing tissue in

frontal and temporal regions[18], and has even been shown to increase volume and thus reverse age-related grey matter loss in the hippocampus[32]. Higher aerobic fitness is also predictive of white matter integrity, in the corpus callosum[33] and frontal and temporal lobes[19]. These beneficial effects on the brain's structural integrity have been linked to cognitive improvements in memory functioning[19,32]. A few studies have investigated the influence of aerobic fitness on functional neural activation, finding a relationship between activation in task-relevant brain regions and fitness during a processing speed task[13] and during attention processing[34]. Despite the clear progress that has thus been made in elucidating the benefits of aerobic fitness for brain structure and function, the complex relationship between brain structure, brain function and cognition in this context still remains poorly understood. More research is needed to study the relationship between vascular functioning, neural functioning, brain structure and cognition, including for language as a crucial domain of cognition.

In the discussion of the literature above, the findings from cross-sectional research and intervention-type research led to converging conclusions with regard the influence of aerobic fitness on cognition, brain structure and brain function. It must be noted that the present research is cross-sectional in nature and thus no causal conclusions can be drawn. Future research will have to determine whether an exercise intervention can successfully increase language abilities.

In summary, in this paper we find an important relationship between age-related decline in language production abilities and aerobic fitness. Word production is an essential step in successful and fluent language production, and the disruption of this process in healthy ageing is detrimental for a positive ageing experience. The present results suggest that higher aerobic fitness are associated with better word production skills in healthy older adults, and thus support the promotion of increased physical activity for healthy ageing and optimal brain function across the life span.

Methods

Participants

53 older adults (34 women, mean age: 70.9 years, SD: 4.4; 19 men, mean age 69.4, SD: 4.9) volunteered to participate in the study. These participants all completed health screening to minimise the risk of an adverse event occurring during the exercise test. Screening information was reviewed by a cardiologist, resulting in 25 older adults being excluded from participating in aerobic fitness testing (for details on exclusion criteria, see below). Consequently, 28 older adults (20 women, mean age: 70.3 years, SD: 4.4; 8 men, mean age: 67.6 years, SD: 5.0) completed the aerobic fitness test and the language experiment. The average education level of the 28 older adults was 16.4 years (SD: 3.2) of formal education, which in the UK starts at 4 years old. The average height was 165.4 cm (SD: 10.5) and the average weight was 66.9 kg (SD: 10.7). For all but 3 older adults we obtained a MOCA score (Montreal Cognitive Assessment) and they scored 26 or higher, which is considered normal.

To provide a baseline against which to compare the older adults' language abilities, 27 young participants (19 women, mean age: 23.4, SD:3.9; 8 men, mean age: 22.9, SD: 2.5) completed

the language experiment. The young participants did not complete aerobic fitness testing. All young participants were currently enrolled as university students with the University of Birmingham.

All participants gave informed consent and were monetarily compensated for participation. All were non-bilingual native British English speakers with no speech or language disorders and no dyslexia. The research was conducted at the University of Birmingham. The research had full ethical approval (UoB ERN_16-0230) and all experimentation was performed in accordance with the relevant guidelines and regulations.

Electrocardiogram (ECG) and general health screening

53 older adults underwent a pre-exercise evaluation prior to the exercise testing. This evaluation consisted of a general health questionnaire, a resting 12-lead electrocardiogram (ECG) assessment and a resting blood pressure measurement, which was reviewed by a cardiologist (MR). Participants who revealed a contraindication to non-medically supervised exercise testing in the general health questionnaire (N=5; e.g., heart condition, family history of heart attack, asthma, prevention medication for stroke), had high resting blood pressure (N=3; systolic >160, diastolic >90), or showed ECG abnormalities (N=12; e.g., S-T suppression, multiple ectopic beats in a row (i.e., >3)) were excluded from the aerobic fitness testing (and referred on to their GP).

Furthermore, for 3 older adults who passed the screening protocol, the exercise testing on the ergometer had to be terminated before a fitness score was obtained, because the participants experienced knee pain. Two older adults choose to withdraw participation post-screening. As a result, fitness scores were obtained for 28 older adults.

Aerobic fitness testing

After screening and inclusion into the study, participants completed a graded sub-maximal aerobic fitness test on a cycle ergometer to estimate maximal oxygen consumption ($V\square O_{2max}$). The sub-maximal fitness test was based on the Åstrand-Rhyming Cycle Ergometer Test, which has been shown to provide a reliable and valid estimate of $V\square O_{2max}$ [35]. Submaximal estimation of maximal aerobic power is a standard procedure for measurement of fitness in sedentary older adults and clinical populations.

For this test participants were asked to cycle on an electromagnetically braked cycle ergometer at 60-70 rpm (rotations per minute). The initial workload began at 35 Watts and then depending on the participant's sex, body mass and habitual physical activity levels, workload increased by 20 to 35 Watt increments every three minutes. This continued until heart rate reached 80% of the participant's estimated maximum heart rate (i.e. 220 minus the participant's age), unless the participant was unable to maintain over 50 rpm or until the participant reached volitional exhaustion. Respiratory gases and volume were collected for measurement of the rate of oxygen consumption ($V\square O_2$). Maximal $V\square O_2$ was then estimated from the relationship between oxygen uptake and heart rate at multiple measurements. The

resulting regression equation predicted participant's $V\square O_{2max}$ (as per the standard procedure: Guiney, et al. [28], Siconolfi, et al. [35]). Prior to this test, participants were asked to abstain from heavy physical exercise and alcohol for 24 hours. They were also instructed not to consume food for 2 hours prior to reporting to the laboratory.

For female participants, the average predicted $V\square O_{2max}$ score was 23.32 (SD = 7.04) with values ranging from 9.4 to 35.1. For male participants, the average predicted $V\square O_{2max}$ score was 31.16 (SD = 6.55) with values ranging from 24.7 to 44.1. It is a standard finding that males have higher $V\square O_{2max}$ scores than females [36]. In general, males have larger body mass (including lung size and cardiovascular capacity) than females, so direct comparison of the raw $V\square O_{2max}$ score for a male and a female is not valid [37]. Scores therefore must be normed or standardized. We calculated z-scores within each sex group for the purpose of relating $V\square O_{2max}$ scores to tip-of-the-tongue occurrence. Using the standardized score allows male and female participants to be viewed on one and the same dimension with regard to $V\square O_{2max}$ scores.

Tip-of-the-tongue experiment

Participants completed a definition filling task: a definition appeared on screen, and participants were asked to indicate whether they knew the word (No/Yes, produce the word) or had a tip-of-the-tongue experience.

The definition materials consisted of 20 definitions of low frequency words (adapted from Jones [38]), 20 questions about people famous in the UK, such as authors, politicians and actors (some adapted from [39]), and 20 definitions of easy words – see Table 3 for examples. Each participant received the 60 definitions in a random order.

The sequence of events on each trial was as follows. A warning signal was displayed for 500 ms after which a definition appeared centred on the screen. The definition remained on screen until the participant responded as follows: they knew the word (button press 'Yes', and then said the word out loud), did not know the word (button press 'No'), or had a tip-of-the-tongue experience (button press 'ToT'). In the instructions to the participants we defined a tip-of-the-tongue experience as: *"Usually we are sure if we know or don't know a word. However, sometimes we feel sure we know a word but are unable to think of it. This is known as a 'tip-of-the-tongue' experience"*.

If participants indicated they experienced a tip-of-the-tongue state, they were asked to provide three pieces of information about its sound structure in response to prompts on the screen which asked them to: 1) guess the initial letter or sound; 2) guess the final letter or sound, and 3) guess the number of syllables. Finally, in order to determine if they were correct in thinking that they knew the target word, participants were asked to select it from a list of four words that were displayed on the screen (the correct answer and three foils – see Table 3 for examples) or to indicate that the word they were thinking of was not in the list.

Table 3: Examples of definitions, target words and foils for multiple-choice questions if participants indicated to have experienced a tip-of-the-tongue.

Definitions	Correct answer	Foils
1. A young goose	Gosling	cygnet, leveret, gelding
2. Able to read and write	Literate	laconic, loquacious, urbane
3. What is the original last name of the boxer who became known as Mohammed Ali?	Clay	Grey, Reid, Grant
4. What was Princess Diana's maiden name?	Spencer	Ogilvy, Lawrence, Philips
5. Ancient tomb for Egyptian kings	Pyramid	crypt, grave, sphynx
6. A fruit of the oak tree eaten by squirrels	Acorn	nut, pit, seed

Data analyses

We analysed the data using mixed effects models, which are an extension of regression models. Mixed effects models are the most suitable models to analyse the present dataset because they can account for the fact that there are repeated observations for both items and participants. We modelled tip-of-the-tongue occurrence using mixed effects logistic regression[40,41] in R[42].

For a categorical outcome variable such as tip-of-the-tongue occurrence, a logistic regression is much more suited than an ANOVA to model the data[40]. Using ANOVA models when the dependent variable is categorical (e.g., yes/no, counts, percentages) can lead to spurious significance values[40,43,44]. In such instances, regression methods are thus preferred[45]. However, ordinary regression analysis ignores correlation of observations within clusters and treats within cluster observations the same as between cluster observations producing invalid standard errors of the fitted coefficients[46]. Any subsequent analysis based on these standard errors (e.g., hypothesis test) is therefore invalid. The use of mixed effect models allows accounting for the fact that there are repeated observations for both items and participants[40,41] and therefore used frequently in psycholinguistic literature.

In addition to modelling tip-of-the-tongue occurrence, we also fitted a model for phonological access scores. We calculated a phonological access score for each trial on which the participant reported a tip-of-the-tongue: 1 point for listing the correct initial sound, 1 point for the correct final sound and 1 point for the correct number of syllables (resulting in a score between 0 and 3). Phonological access scores were modelled using mixed effects linear regression[41] in R[42], again to account for the fact that there are repeated observations for both items and participants.

The regression models for tip-of-the-tongue occurrence and phonological access scores were based on the following predictors: Number of Phonemes of the Target Word, Number of Syllables of the Target Word, Vocabulary Size (% of items named correctly), Education Level (years of formal education), and Age Group (young vs. older adults) / Age (in years). When modelling tip-of-tongue occurrence in the older adults group, we also included a predictor with the standardized $V\square O_{2max}$ scores (see above).

Continuous variables were centered. Group was deviation coded. During the process of model comparison, we started with a model including all factors and then simplified the model using model comparison for fixed effects in stepwise fashion until a model was reached with the lowest AIC value (Akaike information criterion). When a model with a fully specified random effects structure did not converge, we removed random slopes for items before removing any random slopes for participants (since in researcher-designed experiments the variance for items is usually smaller than for participants).

Sample size estimation and power analysis

No previous study has investigated the effects of aerobic fitness on any aspect of language functioning; previous studies have so far reported effects only of aerobic fitness in other cognitive domains: cognitive control, executive functioning, visuo-spatial memory, learning and processing speed (Colcombe & Kramer, 2003). Thus there is no direct evidence on which we could postulate a hypothesized effect size prior to conducting the present study. Rather than postulating a hypothesized effect size based on indirect evidence, we performed a power analysis simulation study which can be a basis for future work. Indeed, any result of a power calculation depends entirely on the size of the hypothesized effect, for which it is impossible to obtain an accurate estimate until direct evidence from a first study is available The current research may then serve as a benchmark for future studies.

We investigated the power of detecting a non-zero effect of aerobic fitness on the probability of experiencing a tip-of-the-tongue state by means of a simulation study under the statistical model describing the present data. As such, we can, based on informative evidence, investigate the Type II error which depends on both the effect size and the sample size. The simulation study allows us to generate data independent of the data described in the current work. Moreover, simulation studies are an effective way of obtaining a power estimate for complex models[47-49] and the approach has been used in the field of psychology in recent years[50-52]. The results of our simulation will serve as in important indicator for postulating sample size in future studies. Further details are described in the results section.

Data Availability

The stimulus materials and the datasets analysed during the current study are available in the OSF repository, [link will be provided before publication].

References

1 Christensen, K., Doblhammer, G., Rau, R. & Vaupel, J. W. Ageing populations: the challenges ahead. *The lancet* **374**, 1196-1208 (2009).

2 Birmingham Policy Commission Report (2014).

3 Lovelace, E. A. & Twohig, P. T. Healthy older adults' perceptions of their memory functioning and use of mnemonics. *Bulletin of the Psychonomic Society* **28**, 115-118 (1990).

4 Maylor, E. A. Age, blocking and the tip of the tongue state. *British Journal of Psychology* **81**, 123-134 (1990).
5 Ossher, L., Flegal, K. E. & Lustig, C. Everyday memory errors in older adults. *Aging, Neuropsychology, and Cognition* **20**, 220-242 (2013).
6 Salthouse, T. A. & Mandell, A. R. Do age-related increases in tip-of-the-tongue experiences signify episodic memory impairments? *Psychological science* **24**, 2489-2497 (2013).
7 Brysbaert, M., Stevens, M., Mandera, P. & Keuleers, E. How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Frontiers in Psychology* **7**, doi:10.3389/fpsyg.2016.01116 (2016).
8 Levelt, W. J. Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences* **98**, 13464-13471 (2001).
9 James, L. E. & Burke, D. M. Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**, 1378 (2000).
10 Burke, D. M. & Graham, E. R. The neural basis for aging effects on language. *The Handbook of the Neuropsychology of Language, Volume 1&2*, 778-800 (2012).
11 Lucas, S. J., Cotter, J. D., Brassard, P. & Bailey, D. M. High-intensity interval exercise and cerebrovascular health: curiosity, cause, and consequence. *Journal of Cerebral Blood Flow & Metabolism* **35**, 902-911 (2015).
12 Thomas, A. G. *et al.* Multi-modal characterization of rapid anterior hippocampal volume increase associated with aerobic exercise. *Neuroimage* **131**, 162-170 (2016).
13 Rosano, C. *et al.* Psychomotor speed and functional brain MRI 2 years after completing a physical activity treatment. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, glq038 (2010).
14 Colcombe, S. J. & Kramer, A. F. Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychological Science* **14**, 125-130, doi:10.1111/1467-9280.t01-1-01430 (2003).
15 Guiney, H. & Machado, L. Benefits of regular aerobic exercise for executive functioning in healthy populations. *Psychonomic bulletin & review* **20**, 73-86 (2013).
16 Bailey, D. M. *et al.* Elevated aerobic fitness sustained throughout the adult lifespan is associated with improved cerebral hemodynamics. *Stroke* **44**, 3235-3238 (2013).
17 Voss, M. W., Vivar, C., Kramer, A. F. & van Praag, H. Bridging animal and human models of exercise-induced brain plasticity. *Trends in cognitive sciences* **17**, 525-544 (2013).
18 Colcombe, S. J. *et al.* Aerobic exercise training increases brain volume in aging humans. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences* **61**, 1166-1170 (2006).
19 Voss, M. W. *et al.* 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**, 2972-2985 (2013).
20 Menenti, L., Segaert, K. & Hagoort, P. The neuronal infrastructure of speaking. *Brain and Language* **122**, 71-80, doi:10.1016/j.bandl.2012.04.012 (2012).
21 Shafto, M. A., Stamatakis, E. A., Tam, P. P. & Tyler, L. K. Word retrieval failures in old age: the relationship between structure and function. *Journal of Cognitive Neuroscience* **22**, 1530-1540 (2010).
22 Shafto, M. A., Burke, D. M., Stamatakis, E. A., Tam, P. P. & Tyler, L. K. On the tip-of-the-tongue: neural correlates of increased word-finding failures in normal aging. *Journal of cognitive neuroscience* **19**, 2060-2070 (2007).

23 Prince, S. A. *et al.* A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity* **5**, doi:10.1186/1479-5868-5-56 (2008).
24 Green, P. & MacLeod, C. J. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* **7**, 493-498 (2016).
25 Rabaglia, C. D. & Salthouse, T. A. Natural and constrained language production as a function of age and cognitive abilities. *Language and cognitive processes* **26**, 1505-1531 (2011).
26 Burley, C. V., Bailey, D. M., Marley, C. J. & Lucas, S. J. Brain train to combat brain drain; focus on exercise strategies that optimize neuroprotection. *Experimental Physiology* **101**, 1178-1184 (2016).
27 Murrell, C. J. *et al.* Cerebral blood flow and cerebrovascular reactivity at rest and during sub-maximal exercise: effect of age and 12-week exercise training. *Age* **35**, 905-920 (2013).
28 Guiney, H., Lucas, S. J., Cotter, J. D. & Machado, L. Evidence Cerebral Blood-Flow Regulation Mediates Exercise-Cognition Links in Healthy Young Adults. *Neuropsychology* **29**, 1-9, doi:10.1037/neu0000124 (2015).
29 Lucas, S. J. *et al.* Role of brain perfusion and oxygenation in exercise-induced improvements in cognition for young and old participants. *Experimental Gerontology* **47**, 541-551 (2012).
30 Giorgio, A. *et al.* Age-related changes in grey and white matter structure throughout adulthood. *NeuroImage* **51**, 943-951 (2010).
31 Erickson, K. I., Leckie, R. L. & Weinstein, A. M. Physical activity, fitness, and gray matter volume. *Neurobiology of aging* **35**, S20-S28 (2014).
32 Erickson, K. I. *et al.* Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences* **108**, 3017-3022 (2011).
33 Johnson, N. F., Kim, C., Clasey, J. L., Bailey, A. & Gold, B. T. Cardiorespiratory fitness is positively correlated with cerebral white matter integrity in healthy seniors. *Neuroimage* **59**, 1514-1523 (2012).
34 Colcombe, S. J. *et al.* Cardiovascular fitness, cortical plasticity, and aging. *Proceedings of the National academy of Sciences of the United States of America* **101**, 3316-3321 (2004).
35 Siconolfi, S. F., Cullinane, E. M., Carleton, R. A. & Thompson, P. D. Assessing VO2max in epidemiologic studies: modification of the Astrand-Rhyming test. *Medicine and science in sports and exercise* **14**, 335-338 (1981).
36 Loe, H., Steinshamn, S. & Wisløff, U. Cardio-respiratory reference data in 4631 healthy men and women 20-90 years: the HUNT 3 fitness study. *PloS one* **9**, e113884 (2014).
37 Heyward, V. H. & Gibson, A. *Advanced fitness assessment and exercise prescription 7th edition*. (Human kinetics, 2014).
38 Jones, G. V. Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition* **17**, 69-76 (1989).
39 Burke, D. M., MacKay, D. G., Worthley, J. S. & Wade, E. On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of memory and language* **30**, 542-579 (1991).
40 Jaeger, T. F. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* **59**, 434-446, doi:10.1016/j.jml.2007.11.007 (2008).

41 Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 255-278 (2013).
42 Team, R. *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL*, <http://www.rstudio.com/.> (2015).
43 Dixon, P. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* **59**, 447-456 (2008).
44 Song, Y., Nathoo, F. S. & Masson, M. E. A Bayesian approach to the mixed-effects analysis of accuracy data in repeated-measures designs. *Journal of Memory and Language* **96**, 78-92 (2017).
45 Quené, H. & Van den Bergh, H. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* **59**, 413-425 (2008).
46 Agresti, A. & Kateri, M. in *International encyclopedia of statistical science* 206-208 (Springer, 2011).
47 Arnold, B. F., Hogan, D. R., Colford, J. M. & Hubbard, A. E. Simulation methods to estimate design power: an overview for applied research. *BMC medical research methodology* **11**, 94 (2011).
48 Johnson, P. C., Barry, S. J., Ferguson, H. M. & Müller, P. Power analysis for generalized linear mixed models in ecology and evolution. *Methods in ecology and evolution* **6**, 133-142 (2015).
49 Bolker, B. M. *Ecological models and data in R.* (Princeton University Press, 2008).
50 Perkovic, S. & Orquin, J. Implicit Statistical Learning in Real World Environments Behind Ecologically Rational Decision Making. *Psychological Science* (2017).
51 Résibois, M. *et al.* The relation between rumination and temporal features of emotion intensity. *Cognition and Emotion*, 1-16 (2017).
52 Doherty, B. R., Patai, E. Z., Duta, M., Nobre, A. C. & Scerif, G. The functional consequences of social distraction: Attention and memory for complex scenes. *Cognition* **158**, 215-223 (2017).

Acknowledgments

This research was funded by a Wellcome Trust ISSF Award to KS, SL and LW. We would like to thank Prof. Janet Lord for helpful feedback when we were conceiving this research. We would also like to thank Denise Clissett for help with participant recruitment, Beth Skinner for her assistance collecting the ECG data, Dr. Evelien Heyselaar for much appreciated advice on the data analyses, and Prof. Ole Jensen for editing the manuscript.

Author contributions

K.S, S.J.E.L. and L.W. designed the study. C.V.B. collected the ECG and aerobic fitness data. A.E.M. collected the language data. M.R. was responsible for the medical screening. P.S. performed power simulations. K.S. analysed the data and wrote the manuscript. All authors edited the manuscript.

Competing interests

The authors declare no competing financial interests.