Enhancing dentate gyrus function with dietary flavanols improves cognition in older adults

Adam M Brickman[1,2,8], Usman A Khan[1,2,8], Frank A Provenzano[1,2,8], Lok-Kin Yeung[1,2], Wendy Suzuki[3], Hagen Schroeter[4], Melanie Wall[5,6], Richard P Sloan[5,6] & Scott A Small[1,2,5,7]

The dentate gyrus (DG) is a region in the hippocampal formation whose function declines in association with human aging and is therefore considered to be a possible source of age-related memory decline. Causal evidence is needed, however, to show that DG-associated memory decline in otherwise healthy elders can be improved by interventions that enhance DG function. We addressed this issue by first using a high-resolution variant of functional magnetic resonance imaging (fMRI) to map the precise site of age-related DG dysfunction and to develop a cognitive task whose function localized to this anatomical site. Then, in a controlled randomized trial, we applied these tools to study healthy 50–69-year-old subjects who consumed either a high or low cocoa–containing diet for 3 months. A high-flavanol intervention was found to enhance DG function, as measured by fMRI and by cognitive testing. Our findings establish that DG dysfunction is a driver of age-related cognitive decline and suggest non-pharmacological means for its amelioration.

The prefrontal cortex[1] and the hippocampal formation[2] are two brain areas that are affected by aging and are thought to contribute to age-related cognitive decline[3]. The hippocampal formation itself is a circuit, which, in its transverse axis, is made up of interconnected regions[4]. Each region houses a population of molecularly and functionally distinct neurons, which accounts for why individual regions are differentially affected by aging and disease[4]. In humans[5,6], non-human primates[7,8] and rodents[7,9,10], the DG is the region of the hippocampal circuit that has shown the most consistent age-related changes using various indicators of functional integrity. Nevertheless, because these observations are only correlational, questions remain as to whether DG dysfunction actually drives age-related memory decline and whether an intervention that improves DG function will ameliorate memory decline in older subjects.

Two recent studies in mice have addressed these questions by showing that a selective enhancement of DG function improved cognitive performance in aging mice, with one study[10] using genetic manipulations and the second[11] using heterochronic parabiosis. In considering more amenable interventions in humans, we were intrigued by a separate study in mice that showed that oral consumption of epicatechin, a dietary flavanol, resulted in a DG-selective increase in dendritic spine density and that the effect was enhanced when the flavanol was administered in conjunction with aerobic exercise[12]. Dendritic function and structure typically covaries with regional metabolism and capillary density[13,14], and, in the same study, the flavanol-induced changes in dendritic spines were found to occur in association with a DG-selective increase in capillary density. An increase in metabolism and capillary density increases cerebral blood volume (CBV), defined as the percentage of a unit tissue comprising vessels. CBV can be mapped *in vivo* with MRI[15], and because it is an established correlate of neuronal metabolism, it is one of the hemodynamic variables used in fMRI[16]. Indeed, we have previously used a high-resolution variant of CBV-fMRI to map exercise-induced changes in both human and mouse DG, which we assumed reflected increases in DG capillary density[17].

Guided by these studies, we hypothesized that a high flavanol dietary intervention, perhaps together with exercise[17,18], would enhance DG function in older humans. First, however, in preparation for an intervention study, we needed to optimize tools with which age-related DG dysfunction can be assessed. Bilaterally, the hippocampal circuit spans the long axis of the medial temporal lobe, from its posterior to its anterior extent. Regions in the circuit, including the DG, extend the long axis and studies have suggested a tripartite functional organization[19], with the long axis typically divided into the head, the body and the tail. In previous studies, we used CBV-fMRI to identify age-related DG dysfunction and to distinguish it from hippocampal patterns of dysfunction associated with Alzheimer's disease (AD)[7,9,20]. The image processing tools used in those studies, however, prohibited an analysis of the full extent of the hippocampal circuit over its long axis. To address this limitation, we recently incorporated new image-processing tools that isolate CBV maps of the hippocampal formation from individual subjects[21] and co-registered them into a group-wise template. In this way, a group-wise voxel-based analysis, in contrast with a single-subject region of interest analysis, can be performed to

[1]Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, New York, USA. [2]Department of Neurology, Columbia University, New York, New York, USA. [3]Center for Neural Science, New York University, New York, New York, USA. [4]MARS Inc., McLean, Virginia, USA. [5]Department of Psychiatry, Columbia University, New York, New York, USA. [6]New York State Psychiatric Institute, New York, New York, USA. [7]Department of Radiology, Columbia University, New York, New York, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to S.A.S. (sas68@columbia.edu).

map the precise location of changes throughout the bilateral hippocampal circuit. In a first series of studies, we applied these neuroimaging tools to 21–65-year-old healthy subjects, establishing the precise pattern of age-related DG dysfunction, and contrasted it with the pattern observed in preclinical stages of AD.

With the growing appreciation that each region of the hippocampal circuit has a distinct cognitive operation[4], an optimized cognitive task is also needed in addition to neuroimaging tools, one that overlaps with the anatomical site of age-related DG dysfunction. Although conventional and validated neuropsychological tests have been designed to evaluate the global hippocampal circuit, we recently found that performance in one of these tasks, the Benton Visual Retention Test (BVRT), localized to the DG among healthy elders[22]. Developed over 50 years ago and ecologically validated[23], the BVRT has a number of practical limitations that are problematic for an intervention study in healthy subjects. First, the test items on the BVRT are not sufficiently challenging for healthy subjects, leading to a 'ceiling' effect' and a non-normal performance distribution. Second, because there are only a limited number of test items, the BVRT is ill-suited for a repeated-measures design. Accordingly, over the last few years we have developed a computerized modification of the BVRT designed to overcome these limitations. We call this task the Modified-Benton (ModBent).

Besides improving its practical utility, to assure that the ModBent is a hippocampal-dependent memory task, one that is particularly sensitive to the aging DG, we informed task design by recent insight. In designing the items of the task, we were guided by the observation that the DG is particularly engaged in the pattern separation of visually similar objects[24]. Procedurally the task was designed to match novel object recognition tasks, which have been validated to localize to the hippocampal circuit[25], and performance in these type of tasks is reliably affected by normal aging in humans[26,27] and other mammalian species[28,29]. Moreover, extending on previous observations[30], a recent study established that performance on novel object recognition depends on the DG and is a cognitive feature of the aging hippocampal circuit[10]. In a second series of studies, we used the ModBent, validated its parametric advantages, confirmed that task performance declined with normal aging and found that it localized to the anatomical site of age-related DG dysfunction. Furthermore, in anticipation of the intervention study, whose goal was to test for a regionally selective effect, we confirmed previous studies showing that a delayed retention task selectively localizes to the entorhinal cortex (EC).

Equipped with these tools, we designed a controlled, randomized trial to investigate in healthy elderly subjects the effects of specific lifestyle modifications, namely dietary cocoa flavanol intake and aerobic exercise, on cognitive and neuroimaging measures of DG function. Although we used an exercise regimen that we have previously shown can improve aerobic fitness in younger subjects[17], the exercise intervention was unexpectedly deemed ineffectual in enhancing aerobic fitness in older subjects. Nevertheless, the cocoa flavanol intervention proved successful, allowing us to test the hypothesis that age-related DG dysfunction contributes to cognitive aging.

RESULTS

Mapping age-related DG dysfunction across the hippocampal circuit

As previously described[31], we generated high-resolution CBV brain maps using gadolinium-enhanced T1-weighted scans acquired perpendicular to the hippocampal long-axis, with sub-millimeter in-plane resolution of 0.68×0.68 mm and slice thickness of 3 mm (O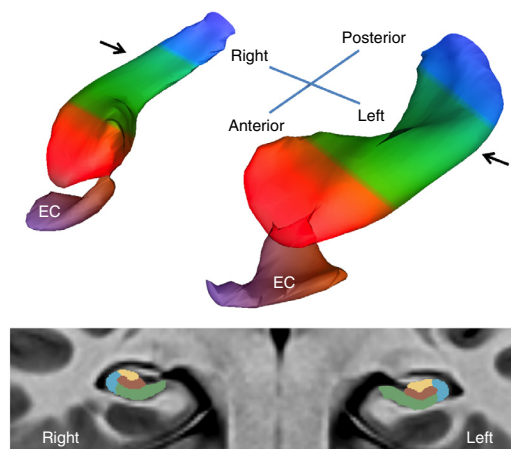nline Methods). To perform group-wise voxel-based analyses on the whole hippocampal circuit, we first isolated hippocampal subfields using FreeSurfer segmentations[32] and thresholded posterior probabilities to generate a composite mask of the bilateral hippocampal formation. These masked T1-weighted images were then used to generate a group-wise template to which individual images were co-registered using a diffeomorphic technique that maximizes cross-correlation among images[33].

To map the precise localization of age-related changes in the hippocampal circuit, we applied this processing approach to CBV scans acquired from 35 healthy individuals, 21–65 years of age, and generated a three-dimensional rendering from the group-wise template. The approach generated a high-fidelity rendering of the bilateral hippocampal circuit throughout its long axis, whereas the sub-millimeter resolution of the native CBV scans preserved the regional anatomy in its transverse axis (**Fig. 1** and **Supplementary Video 1**). The rendering allowed the results of voxel-based analyses to be displayed in a single snapshot of the hippocampal circuit. Replicating prior studies[5,7,9], the DG was found to be the region most reliably affected in association with aging (**Fig. 2a**). Extending prior observations[5,7,9], we observed the greatest age-related DG changes in the body of the hippocampus.

Because AD is a disorder that also targets the hippocampal circuit of aging individuals, we contrasted the anatomical pattern of age-related hippocampal dysfunction to that of AD by applying the same processing tools to a previously published data set in which CBV-fMRI was generated in patients in the earliest preclinical stages of AD[21]. Using the same processing and rendering tools, we observed an anatomical dissociation with aging (**Fig. 2b**) whereby, early on, AD differentially targets the EC.

A cognitive task that localizes to age-related DG dysfunction

The stimuli used in the ModBent task were 'Lissajous' figures derived by parametrically manipulating equations that generate intersecting
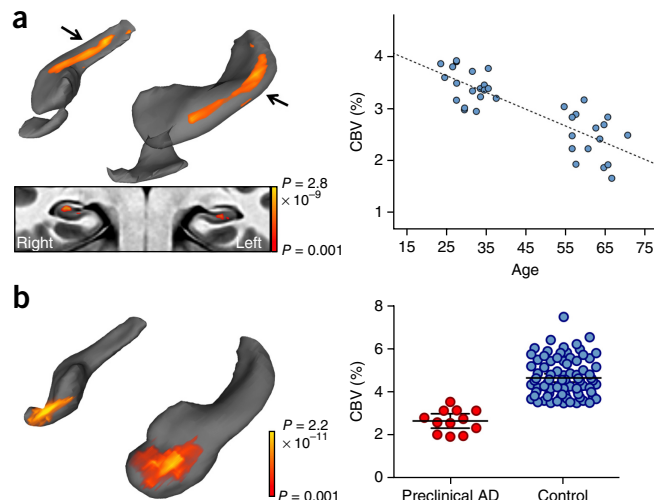

Figure 1 A bilateral map of the hippocampal circuit generated from the high-resolution acquisitions of CBV-fMRI. A three-dimensional rendering of the bilateral hippocampal circuit (top) derived from a group-wise template of multiple axial slices (illustrated at bottom), generated using the native sub-millimeter resolution of CBV maps (**Supplementary Video 1**). The EC is the main gateway into the hippocampal circuit, and over the long axis (top) the circuit is divided into the head (red), body (green) and tail (blue). In its transverse axis (bottom), taken through the body of the hippocampal long axis (indicated by arrows, top), the hippocampal circuit is divided into the dentate gyrus (brown), CA3 (yellow), CA1 (blue) and subiculum (green). All statistical analyses were performed only in the boundaries of the hippocampal circuit.

Figure 2 Mapping a differential pattern of age-related dysfunction in the hippocampal circuit. (**a**) A voxel-based analysis of 35 individuals ranging from 21–65 years of age revealed that the greatest age-related decline in CBV occurred in the body of the hippocampal circuit (top left; color coded by degree of significance). A transverse slice (bottom left), onto which the hippocampal circuit mask was applied, revealed that age-related CBV decline localized primarily to the dentate gyrus. A scatter plot (right) shows the association between age and mean CBV from all significantly correlated voxels ($\beta = -0.844$, $r^2 = 0.678$, $P < 0.001$). (**b**) A voxel-based analysis in subjects with preclinical Alzheimer's disease compared with age-matched controls revealed CBV reductions in the EC (left, reprocessed from ref. 1). A scatter plot (right) shows individual-subject mean CBV values in those lateral EC voxels determined to be significantly different between patients with preclinical Alzheimer's disease and healthy controls ($t_{94} = 7.265$, $P < 0.001$).

sinusoidal curves (**Supplementary Fig. 1**). Informed by studies that established that the DG is involved in pattern separation[24], our approach to stimulus generation allowed us to create stimuli that were very similar to each other in a mathematically controlled manner. The procedure of the ModBent was designed as a novel object recognition task, guided by a recent study that established that this type of task is a cognitive feature of age-related hippocampal dysfunction that localizes to the DG[10]. During the immediate matching trials, participants viewed a single complex stimulus for 10 s; following a 1-s delay, they were asked to select which one of two objects was identical to the studied stimulus. Following 41 matching trials, participants were shown serially individual complex objects and asked to indicate whether each object was identical to any of the target stimuli studied during the immediate matching trials. There were 82 recognition trials, which included 41 targets and 41 foils (**Fig. 3a**). The primary dependent variable for the ModBent is the mean reaction time for correct rejections of foil stimuli on the delayed recognition trials. Two versions of the ModBent were created with the psychological software ePrime and administered on a standard laptop computer.

To test and validate the task, we applied the ModBent to 62 undergraduate students (mean age = 21.13 ± 0.70 years; **Supplementary Table 1**). All subjects were able to understand the instructions
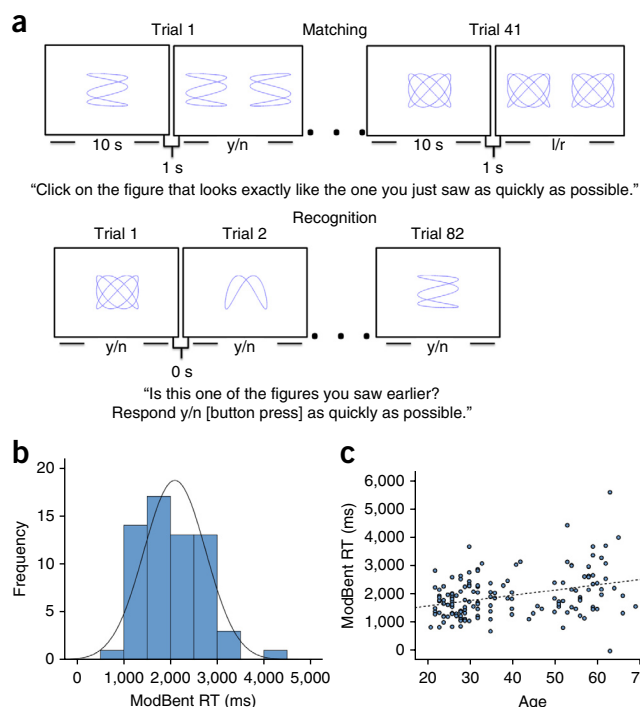
and performance was normally distributed (**Fig. 3b**). We tested 12 participants following a 3-month interval; the intra-class coefficient (ICC) was 0.743, which indicates an acceptable test–re-test reliability for the scale[34]. Next, we applied the ModBent to 149 healthy subjects across the adult lifespan, from 21–69 years of age (**Supplementary Table 1**), and found that ModBent performance worsened with age (**Fig. 3c**) at a rate of approximately 220 ms per decade ($\beta = 22.31$, $P < 0.001$).
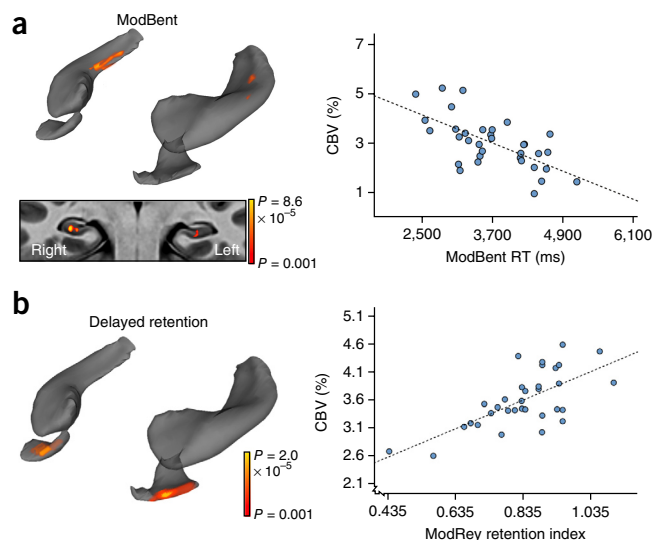
To determine whether performance on the ModBent localizes to the DG, we applied the ModBent to 35 young healthy subjects (**Supplementary Table 1**), ranging in age from 22 to 41, who were also imaged with CBV-fMRI. Better performance on the ModBent correlated selectively with higher DG CBV in the body of the hippocampus (**Fig. 4a**). We also administered a delayed retention task to the 35 subjects, as previous studies have suggested that this cognitive operation localizes to the EC. Confirming previous observations[22], performance on delayed retention was found to correlate selectively with EC CBV (**Fig. 4b**), thereby establishing an anatomical double dissociation with the ModBent.

High flavanol intake improves DG function

Armed with these imaging and cognitive tools, we conducted a controlled randomized, double-masked, 3-month dietary intervention study[35] (**Supplementary Fig. 2**), during which healthy, but sedentary,

Figure 3 Performance on the ModBent declines with age. (**a**) The ModBent task is divided into two parts. During the matching trials (top), participants were shown a complex stimulus for 10 s. Following a 1-s inter-trial interval, they were required to select, via a key press, which of two stimuli matched the one they had just studied, as quickly and as accurately as possible. Following 41 matching trials, participants were shown a series of 82 stimuli (bottom), 41 of which appeared on the initial study set during the matching trials and 41 of which were foils. They were required to indicate, as quickly as possible, whether each stimulus appeared earlier (yes response) or was new (no response). The primary variable of the ModBent task is the mean RT for correct rejections of foils during the recognition trials. (**b**) As part of the task development, the ModBent was administered to 62 undergraduate students (mean age = 21.1 ± 0.80 years). Performance was normally distributed (Kolmogorov-Smirnov test statistic (62) = 0.091, $P = 0.20$), as shown. A subset of participants were tested following a 3-month interval and good test–re-test reliability was achieved (ICC = 0.743). (**c**) When applied to 149 healthy subjects across the adult life span (ages 21–69 years), performance on the ModBent (reaction time in correctly rejecting foil items) worsened with age ($\beta = 22.31$, $P < 0.001$).

a ModBent

b Delayed retention

Figure 4 Performance on the ModBent overlaps with the anatomical site of hippocampal aging. (**a**) A voxel-based analysis of 35 individuals revealed that better performance on the ModBent task correlated with greater CBV in the body of the hippocampal circuit (top left). A transverse slice, onto which the hippocampal circuit mask was applied, shows that the correlation localized primarily to the dentate gyrus (bottom left; coronal section, color coded by degree of significance). A scatter plot (right) shows the association between ModBent performance and single-subject mean CBV from all significantly correlated voxels ($\beta = -0.633$, $r^2 = 0.465$, $P < 0.001$). (**b**) A voxel-based analysis in the same group showed that better performance on a delayed retention task correlated with greater CBV in the EC (left). A scatter plot (right) shows the association between performance on this delayed retention task and single-subject mean CBV from all significantly correlated voxels ($\beta = 0.693$, $r^2 = 0.562$, $P < 0.001$).

50–69-year-old subjects were randomly assigned to one of four experimental groups (**Table 1**): a high-flavanol intake group (900 mg cocoa flavanols and 138 mg of (−)-epicatechin per d) with or without a regimen of aerobic exercise, and a low-flavanol intake group (10 mg cocoa flavanols and <2 mg (−)-epicatechin per d) with or without a regimen of aerobic exercise (**Supplementary Table 2**).

As in our previous study in young subjects[17], the exercise regimen consisted of 1 h per d of aerobic exercise, 4 d per week. All subjects were imaged with CBV-fMRI and were evaluated with the ModBent at baseline, and then a second time at the end of the study using a different set of stimuli. In addition to hippocampal CBV and ModBent as the primary outcome measures, aerobic capacity (VO$_2$ max) was included as a secondary measure to determine whether the aerobic fitness regimen was effective. In addition, all subjects were evaluated with the delayed retention task to test for the specificity of any observed effect with the primary behavioral outcome.

A total of 37 subjects completed the study, with an equivalent distribution across the four experimental conditions (**Table 1**). There were no group differences for age ($F_{3,33} = 0.547$, $P = 0.653$), education ($F_{3,33} = 0.413$, $P = 0.744$) or sex ($X^2 = 0.736$, $P = 0.865$). An ANCOVA revealed that a high-flavanol intervention had a significant effect on ModBent performance, independent of exercise (high-flavanol ModBent reaction time (RT) = 1,997 ms, low-flavanol ModBent RT = 2,627 ms, $t_{31} = 2.17$, $P = 0.038$, Cohen's $d = 0.816$), but had no effect on delayed retention (high-flavanol delayed retention = 0.77, low-flavanol delayed retention = 0.78, $t_{29} = 0.19$, $P = 0.85$, Cohen's $d = 0.06$). Compared with the low-flavanol intervention, subjects who received the high-flavanol intervention showed a mean improved cognitive performance of 630 ms (**Supplementary Fig. 3**). The
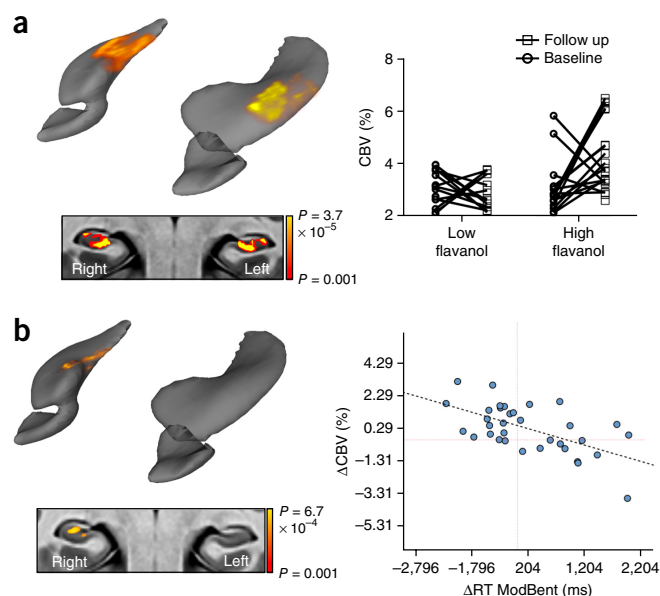
ANCOVA was designed as a between-group comparison controlling for each individual's baseline performance.

No significant effect was observed for exercise intervention on the ModBent (exercise group ModBent RT = 2,346 ms, control group ModBent RT = 2,277 ms, $t_{31} = 0.24$, $P = 0.815$, Cohen's $d = 0.09$) or on delayed retention (exercise group delayed retention = 0.795, control group delay retention = 0.757, $t_{29} = -0.56$, $P = 0.581$, Cohen's $d = 0.187$). A secondary analysis revealed that exercise unexpectedly had no effect on VO$_2$ max (exercise group VO$_2$ max = 26.7, control

a

b

Figure 5 Flavanols enhance CBV-fMRI. (**a**) A voxel-based analysis of 37 individuals revealed that, compared with those in the low-flavanol condition, those in the high-flavanol condition showed a CBV increase in the body of the hippocampal circuit. A transverse slice, onto which the hippocampal circuit mask was applied, shows that the CBV increase localized primarily to the dentate gyrus and subiculum (bottom left; color coded for degree of significance). Single-subject line plots (right) show mean CBV in all voxels determined to have a significant group × time interaction ($F_{1,33} = 27.58$, $P < 0.0001$). (**b**) A voxel-based analysis of 35 individuals revealed that better performance on the ModBent task correlated with greater CBV in the body of the right hippocampal circuit (top left). A transverse slice, onto which the hippocampal circuit mask was applied, shows that the correlation localized primarily to the dentate gyrus (bottom left; color coded for degree of significance). A scatter plot (right) shows the association between change in ModBent performance and single-subject mean change in CBV from all significantly correlated voxels ($\beta = -0.554$, $r^2 = 0.307$, $P < 0.001$).

Table 1 Distribution and demographics of subjects randomized to the flavanol and exercise intervention study

	Exercise	No exercise
High flavanol	$N = 8$	$N = 11$
	Mean age = 57 years	Mean age = 57 years
	Mean education = 19 years	Education = 19 years
	Percentage of women = 75	Percentage of women = 64
Low flavanol	$N = 9$	$N = 9$
	Mean age = 59 years	Mean age = 58 years
	Mean education = 19 years	Mean education = 17 years
	Percentage of women = 78	Percentage of women = 78

group VO_2 max = 25.5, t_{31} = 1.21, P = 0.237, Cohen's d = 0.28) and the exercise intervention was therefore deemed a treatment failure.

The imaging data were analyzed with a repeated-measures ANOVA model that examined a group (high versus low flavanol) by time interaction. Subjects in the high-flavanol condition showed a significant increase of CBV in the DG and the downstream subiculum in the body on the hippocampus (group × time interaction: $F_{1,33}$ = 27.58, $P < 0.0001$, η^2 = 0.471; high-flavanol effect: $F_{1,33}$ = 4.487, P = 0.042, η^2 = 0.124; high-flavanol effect: $CBV_{baseline}$ = 2.9 ± 1.05, $CBV_{followup}$ = 4.7 ± 1.42; low-flavanol effect: $CBV_{baseline}$ = 3.1 ± 0.64, $CBV_{followup}$ = 2.8 ± 0.68; **Fig. 5a**). Finally, we performed a correlation between changes in the ModBent and changes in CBV. The changes in the ModBent were associated with CBV changes in the body of the DG (**Fig. 5b**). Consistent with our behavioral findings, the dietary intervention had no effect on the EC.

DISCUSSION

Previous studies have demonstrated that flavanol consumption can enhance blood flow[36,37], whereas other studies have shown that flavanol consumption can increase dendritic spine and capillary density in the DG[12]. Blood flow and capillary density are interlinked and both will influence CBV. The observed increase in DG function induced by cocoa flavanol intake might be caused by a more global increase in blood flow or it might be caused by a more selective increase in capillary density. We favor the latter interpretation because of the confined anatomical distribution of the observed effect. By either mechanism, however, our results indicate that dietary cocoa flavanol consumption enhanced DG function in the aging human hippocampal circuit.

We note that previous studies[18], including our own[17], have documented that aerobic exercise, when effective at increasing aerobic fitness, typically improves hippocampal function. Why then was the exercise arm of our study an intervention failure, as evidenced by the lack of improvement in aerobic fitness, whereas our previous study documented that an identical exercise regimen was successful in enhancing aerobic fitness in a younger population[17]? One possibility is that a more stringent exercise regimen is required at the older ages that we investigated. In any case, the purpose of our study was not to test whether exercise will improve hippocampal function; we accept as a fact that aerobic exercise, when effective[17,18], will improve the function of the aging hippocampal circuit. The purpose of the study was to test the hypothesis that, if the DG is a source of age-related hippocampal dysfunction, then an intervention that enhances DG function will improve a cognitive phenotype that characterizes the aging hippocampal circuit. This turned out to be case with flavanol consumption.

The ModBent was procedurally designed to be a novel object recognition task that is distinguished by the use of stimuli that were deliberately generated to have high visual similarities, ensuring that performance is heavily dependent on pattern separation. This cognitive loading likely accounts for why the ModBent localized primarily to the DG and provides further support that, as established in aging mice[10], worsening performance in novel object recognition is a characteristic of the aging hippocampal circuit. Improvement on the ModBent in the high-flavanol group was equivalent to improvements in cognition by approximately three decades of life, shifting the slope of age-related memory decline (data not shown), which is qualitatively similar to the normalizing effect on novel object recognition observed in aging mice who received a genetic manipulation delivered to the DG[10]. Together, these results provide evidence that age-related changes in the DG observed in aging humans underlie and drive a hippocampal-dependent component of cognitive aging.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
A.M.B. designed and implemented the ModBent task and help write the manuscript. U.A.K. and F.A.P. performed the imaging analyses and help write the manuscript. L.-K.Y. aided in designing the ModBent task. W.S. administered the ModBent task to college students. H.S. aided is establishing inclusionary/exclusionary criteria for the clinical trial. M.W. performed the statistical analysis on the cognitive variables. R.P.S. was responsible for subject recruitment and characterization and helped to write the manuscript. S.A.S. designed and evaluated all the studies and was the primary writer of the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Gazzaley, A., Cooney, J.W., Rissman, J. & D'Esposito, M. Top-down suppression deficit underlies working memory impairment in normal aging. *Nat. Neurosci.* **8**, 1298–1300 (2005).
2. Small, S.A., Stern, Y., Tang, M. & Mayeux, R. Selective decline in memory function among healthy elderly. *Neurology* **52**, 1392–1396 (1999).
3. Morrison, J.H. & Baxter, M.G. The ageing cortical synapse: hallmarks and implications for cognitive decline. *Nat. Rev. Neurosci.* **13**, 240–250 (2012).
4. Small, S.A., Schobel, S.A., Buxton, R.B., Witter, M.P. & Barnes, C.A. A pathophysiological framework of hippocampal dysfunction in ageing and disease. *Nat. Rev. Neurosci.* **12**, 585–601 (2011).
5. Small, S.A., Tsai, W.Y., DeLaPaz, R., Mayeux, R. & Stern, Y. Imaging hippocampal function across the human life span: is memory decline normal or not? *Ann. Neurol.* **51**, 290–295 (2002).
6. Yassa, M.A., Mattfeld, A.T., Stark, S.M. & Stark, C.E. Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proc. Natl. Acad. Sci. USA* **108**, 8873–8878 (2011).
7. Small, S.A., Chawla, M.K., Buonocore, M., Rapp, P.R. & Barnes, C.A. Imaging correlates of brain function in monkeys and rats isolates a hippocampal subregion differentially vulnerable to aging. *Proc. Natl. Acad. Sci. USA* **101**, 7181–7186 (2004).
8. Hara, Y. *et al.* Synaptic distributions of GluA2 and PKMzeta in the monkey dentate gyrus and their relationships with aging and memory. *J. Neurosci.* **32**, 7336–7344 (2012).
9. Moreno, H. *et al.* Imaging the abeta-related neurotoxicity of Alzheimer disease. *Arch. Neurol.* **64**, 1467–1477 (2007).
10. Pavlopoulos, E. *et al.* Molecular mechanism for age-related memory loss: the histone-binding protein RbAp48. *Sci. Transl. Med.* **5**, 200ra115 (2013).
11. Villeda, S.A. *et al.* Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice. *Nat. Med.* **20**, 659–663 (2014).
12. van Praag, H. *et al.* Plant-derived flavanol (–)epicatechin enhances angiogenesis and retention of spatial memory in mice. *J. Neurosci.* **27**, 5869–5878 (2007).
13. Borowsky, I.W. & Collins, R.C. Metabolic anatomy of brain: a comparison of regional capillary density, glucose metabolism and enzyme activities. *J. Comp. Neurol.* **288**, 401–413 (1989).
14. Fonta, C. & Imbert, M. Vascularization in the primate visual cortex during development. *Cereb. Cortex* **12**, 199–211 (2002).
15. Lin, W., Celik, A. & Paczynski, R.P. Regional cerebral blood volume: a comparison of the dynamic imaging and the steady state methods. *J. Magn. Reson. Imaging* **9**, 44–52 (1999).
16. Belliveau, J.W. *et al.* Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* **254**, 716–719 (1991).
17. Pereira, A.C. *et al.* An *in vivo* correlate of exercise-induced neurogenesis in the adult dentate gyrus. *Proc. Natl. Acad. Sci. USA* **104**, 5638–5643 (2007).

18. Erickson, K.I. *et al.* Exercise training increases size of hippocampus and improves memory. *Proc. Natl. Acad. Sci. USA* **108**, 3017–3022 (2011).

19. Volpe, J.J., Herscovitch, P., Perlman, J.M. & Raichle, M.E. Positron emission tomography in the newborn: extensive impairment of regional cerebral blood flow with intraventricular hemorrhage and hemorrhagic intracerebral involvement. *Pediatrics* **72**, 589–601 (1983).

20. Wu, W. *et al.* The brain in the age of old: the hippocampal formation is differentially affected by diseases of late life. *Ann. Neurol.* **64**, 698–706 (2008).

21. Khan, U.A. *et al.* Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nat. Neurosci.* **17**, 304–311 (2014).

22. Brickman, A.M., Stern, Y. & Small, S.A. Hippocampal subregions differentially associate with standardized memory tests. *Hippocampus* **21**, 923–928 (2011).

23. Strauss, E., Sherman, E.M.S. & Spreen, O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (Oxford University Press, 2006).

24. Yassa, M.A. & Stark, C.E. Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).

25. Broadbent, N.J., Squire, L.R. & Clark, R.E. Spatial memory, recognition memory and the hippocampus. *Proc. Natl. Acad. Sci. USA* **101**, 14515–14520 (2004).

26. Schacter, D.L., Cooper, L.A. & Valdiserri, M. Implicit and explicit memory for novel visual objects in older and younger adults. *Psychol. Aging* **7**, 299–308 (1992).

27. Grady, C.L. *et al.* Age-related reductions in human recognition memory due to impaired encoding. *Science* **269**, 218–221 (1995).

28. Moss, M.B., Rosene, D.L. & Peters, A. Effects of aging on visual recognition memory in the rhesus monkey. *Neurobiol. Aging* **9**, 495–502 (1988).

29. Erickson, C.A. & Barnes, C.A. The neurobiology of memory changes in normal aging. *Exp. Gerontol.* **38**, 61–69 (2003).

30. Jessberger, S. *et al.* Dentate gyrus-specific knockdown of adult neurogenesis impairs spatial and object recognition memory in adult rats. *Learn. Mem.* **16**, 147–154 (2009).

31. Khan, U.A. *et al.* Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nat. Neurosci.* **17**, 304–311 (2014).

32. Van Leemput, K. *et al.* Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* **19**, 549–557 (2009).

33. Avants, B.B., Epstein, C.L., Grossman, M. & Gee, J.C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).

34. Anastasi, A. & Urbina, S. *Psychological Testing* (Prentice Hall, Upper Saddle River, New Jersey, 1997).

35. New York State Psychiatric Institute & Mars, Inc. Mars flavanol exercise and cognitive function study. *ClinicalTrials.Gov* http://clinicaltrials.gov/show/NCT01180127 (2014).

36. Schroeter, H. *et al.* (–)-Epicatechin mediates beneficial effects of flavanol-rich cocoa on vascular function in humans. *Proc. Natl. Acad. Sci. USA* **103**, 1024–1029 (2006).

37. Sorond, F.A., Hollenberg, N.K., Panych, L.P. & Fisher, N.D. Brain blood flow and velocity: correlations between magnetic resonance imaging and transcranial Doppler sonography. *J. Ultrasound Med.* **29**, 1017–1022 (2010).

ONLINE METHODS

CBV-fMRI. We used a steady-state contrast enhancement CBV technique. that was optimized to image the hippocampal circuit. Image acquisition and processing: Images were acquired with a Philips Achieva 3.0-T MRI scanner. A gradient echo T1-weighted scan (TR = 6.7, TE = 3.1, FOV = 240 × 240 × 192 mm, voxel size = 0.9 × 0.9 × 0.9 mm) was acquired before a pair of un-scaled T1-weighted images (TR = 7, TE = 3, FOV = 240 × 240 × 196 mm, voxel size = 0.68 × 0.68 × 3 mm). The image resolution used results from a systematic exploration (data not shown) of the scan protocol's parameters. Care was taken to allow sufficient visualization of the subfields of the hippocampus and reduce volume-averaging artifact and to minimize the countervailing loss of signal detection and increase in signal variability at higher resolutions.

Images were aligned perpendicular to the long-axis of each patient's hippocampus with the aid of a sagittal scout image. Next, the coronal T1-weighted CBV sequence FOV was spatially adjusted to capture transverse slices of the hippocampal formation. Scans were acquired before and after a bolus injection of a gadolinium based contrast agent (Omniscan, GE Healthcare) to produce a CBV snapshot of the subfields of the hippocampal formation throughout its full longitudinal axis.

Each subject's isometric T1-weighed MRI image was processed using FreeSurfer (http://surfer.nmr.mgh.harvard.edu) anatomical segmentation software[32], which generates a whole brain cortical parcellation as well as a segmentation of the subfields of the hippocampus. Specifically, this approach[32] uses a Bayesian inferential framework to parameterize the classification of individual regions (DG/CA4, CA3, CA1, subiculum) as well as other structures (fimbria, hippocampal fissure, choroid plexus and inferior horn of the lateral ventricle). Parameterization of the other structures reduces type I error in classification of the subfields and extends the classic FreeSurfer segmentation, which assigns a general hippocampus label.

To address the issue of type II error, we thresholded individual subfield (DG/CA4, CA3, CA1, subiculum) maps liberally at 50% posterior probability and combined the thresholded maps into a composite mask of the hippocampus. This composite mask, rather than individual subfield masks, which may contain segmentation errors, was then serially dilated by one voxel in the coronal plane to generate a mask of the entire hippocampal formation and surrounding tissue (that is, the medial temporal lobe).

Then, as previously described in greater detail[21], each subject's pre and post-gadolinium images were co-registered to generate a subject-specific median template using a robust, inverse-consistent linear registration[38]. This subject-specific median template was then co-registered to the isometric T1-weighted image space using the same linear registration.

Generating CBV maps. For each individual, pre-contrast images were subtracted from post-contrast images to generate a raw subtracted volume. Intra-subject and intra-modal co-registrations used a symmetric rigid body alignment[38]. Raw image values were normalized to the mean signal intensity of the subject's superior sagittal sinus. As previously described in detail[21], the superior sagittal sinus was isolated in each post-contrast image using a modified Frangi vesselness[39] filter and pre-defined ROI.

Co-registration. Voxel-based analysis was performed separately on the left and right hippocampal formations by first combining and binarizing the following FreeSurfer regional segmentations: DG/CA4, CA2/3, CA1, subiculum, presubiculum, hippocampus and EC[32]. This mask was then transformed into CBV space using a linear transform, which was estimated using a robust rigid registration method[38]. Pre-contrast images, corresponding to each subject's CBV image, were masked using the transformed image to produce a hippocampal-specific grayscale image. Then, as previously described in greater detail[21], images were co-registered into a group-wise template space using a symmetric diffeomorphic algorithm[33]. Nonlinear warps parameterizing single-subject co-registrations were used to transform three images into template space: the source pre-contrast T1-weighted hippocampal formation grayscale volume, its corresponding CBV volume and a new mask of the hippocampal formation. This composite mask was generated using a 90% posterior probability threshold (rather than 50%, previously), which served to reduce type I error in identification of the hippocampal formation in group-wise template space. The transformed masks were then summed and the resulting image was thresholded at 90% overlap to generate a final hippocampal volume mask in template space, which restricted our analysis

to those voxels most likely belonging to the hippocampal formation. Notably, for all analyses, the hippocampal fissure and fimbria were excluded from the final masks as a result of their high vascularization and given that these structures were not of interest.

Transformed CBV maps were analyzed with a general linear model implemented in SPM8 (Wellcome Department of Imaging Neuroscience). For correlational analyses, data were analyzed with a multiple regression model, including sex as a covariate and age or cognitive test performance as regressors. For the intervention study, 35 subjects had complete CBV data at both time points. A factorial ANOVA model was used with 'time' (time 1 versus time 2) as the repeated-measures factor, and 'group' (high- versus low-flavanol supplement) as the between-subject factor, where subject was included as a nested random factor[6]. Sex was included as a covariate. Individual groups were contrasted using Student's t test, and the treatment-time interaction effect was contrasted with an F test.

Results were thresholded at $P < 0.001$ and corrected for multiple comparisons at the cluster level using a Monte-Carlo simulation run for 10,000 iterations to yield a corrected $P < 0.05$. Corrected maps from individual group comparisons and regressions were then overlaid onto their respective group-wise templates in cross-section using 3DSlicer (http://www.slicer.org/), and also displayed as maximum intensity projections over mesh models of the hippocampal formation.

The ModBent task. The stimuli used in the ModBent are Lissajous figures, intersecting sinusoidal curves defined by the parametric equations $X = \sin(at + d)$, $Y = \sin(bt)$. a represents the vertical frequency (that is, how many maxima the sinusoid has vertically), b represents the horizontal frequency, d represents the phase (that is, the "offset" of the stimuli), and t is a variable for the parametric equation used to define the number of stimuli to be generated.

Stimulus sets were generated using unique combinations of a and b (that is each set is defined by the combination of a and b values; **Supplementary Fig. 1**). a values were ranged from [1–8, 11]; a_9 and a_{10} were excluded after pilot analysis showed that they were indistinguishable from one another. For $a = [1–7]$, b values ranged from [1–6]; only those b values were used that produced non-integer quotients when divided by a (that is $b = 6$, $a = 2$ was not used), to exclude simpler iterations of the same stimulus. For $a = [8, 11]$, b values ranged from [2–($a − 1$)]. $b = 1$ was excluded because at $a = [8, 11]$, as it was too small to perceive.

Each a/b combination in a stimulus set also contained five phase-shifted alternatives, which were determined by variable $d = [1–5]$ (**Supplementary Fig. 1**). Stimulus sets with $d = [1–4]$ were used in the matching task, whereas $d = 5$ was reserved for the incorrect response/lures on the recognition task. Each $a/b/d$ stimulus set was partitioned into two groups of stimuli, which were assigned pseudo-randomly, that is group one: 3/2/4, 3/2/1; group two: 3/2/3, 3/2/2. The correct stimulus in matching task trials was determined randomly for each participant. Shapes were drawn as closed figured using values of $t = [1–20\pi]$, sampled at intervals of $\pi/1,000$. This insured that a complete, closed-loop figure would be formed for every a/b pair.

In each immediate matching trial, the pair of stimuli shown had the same vertical and horizontal frequencies (the combination of which is trial unique), but varied in phase. This approach allowed us to create stimuli that were very similar to each other (both within and across trials) and varied systematically across the entire set. During the immediate matching trials, participants viewed a single complex stimulus for 10 s; following a 1-s delay, they were asked to select which one of two objects was identical to the studied stimulus. After 41 matching trials, participants were shown serially individual complex objects and asked to indicate whether the object was identical to any of the target stimuli studied during the immediate matching trials. There are 82 recognition trials, which include 41 targets and 41 foils. The primary dependent variable for the ModBent is the mean reaction time for correct rejections of foil stimuli on the delayed recognition trials. Two versions of the ModBent were created with the psychological software ePrime (Psychology Software Tools) and administered on a standard laptop computer.

The modified Rey auditory learning task. 20 non-semantically or phonemically related words were presented over three learning trials, in which the test administrator read the word list and the subject freely recalled as many words possible. Administration of the three learning trials was immediately followed

by one learning trial of a distracter list and then a short delayed freely recall of the initial list. After a 60-min delay period, subjects were asked to freely recall words from the initial list and then to freely recall items from the distracter list. They were then administered a forced-choice recognition trial in which they were required to identify the 20 words from the initial learning trial among semantically and phonemically related words as well as words from the distracter trial. Finally, a source memory trial was administered in which subjects were read a list containing only words from the initial learning list and from the distracter list and were asked to identify from which each word came. For the purposes of this study, the dependent measure was a retention score for which the number of words recalled after the short delay was divided by the number of words recalled on the third learning trial.

Randomized controlled trial. All subjects provided informed consent. The Institutional Review Board of the New York State Psychiatric Institute approved this study. The trial was registered with ClinicalTrials.Gov (NCT01180127).

Study participants. Study participants were healthy, sedentary older adults, 50–69 years of age, recruited from the Columbia University Medical Center/New York Presbyterian Hospital campus. Participants were eligible if they did not exercise regularly or exceed American Heart Association standards for average fitness (VO_2 max < 36 and 33 ml kg^{-1} min^{-1} for men age 50–59 and 60–69, respectively; < 29 and 27 ml kg^{-1} min^{-1} for women age 50–59 and 60–75, respectively, established by cardiopulmonary exercise testing (CPET)). Exclusion criteria included use of psychotropic medications, current psychiatric disorder, any condition for which aerobic training as counter-indicated, habitual consumers of dietary or herbal supplements including Gingko, flavonoid, and dietary herbal or plant extracts, lactose intolerance, and diabetes.

417 potential participants were screened, of whom 131 initially were determined to be eligible. 88 provided informed consent, 51 met the CPET standards, 46 successfully completed the run-in phase (see below) and 41 were randomized. Participants were randomized as follows: high dietary flavanol (DF) + active exercise (AE) ($n = 10$), high DF + wait list control ($n = 11$), low DF + AE ($n = 10$) and low DF + wait list control ($n = 10$). 37 participants completed the study and 4 dropped out. Subjects receiving high DF were instructed to take 2 450-mg high-flavanol supplements, for a 900-mg daily dose. Subjects receiving low DF consumed a daily total of 45 mg, divided into two doses. There were no other dietary interventions, although we excluded habitual consumers of dietary or herbal supplements.

Study protocol. Participants completed a phone screening for initial determination of eligibility. Those who passed this screen met with a study research assistant to provide informed consent and for further determination of eligibility, after which they completed a cardiopulmonary exercise test. Those who met aerobic capacity inclusion criteria completed a 2-week run-in period during which they went to the fitness center four times per week for 1-h-long stretching sessions. To continue in the study, participants had to attend at least seven of these eight sessions. Qualifying participants then completed a battery of neuropsychological tests and an MRI session. After completing these tests, they were randomized to a treatment condition.

Participants assigned to the AE condition trained in the Columbia University Fitness Center four times per week for 12 weeks. Each exercise session lasted about 1 h: 5 min of low-intensity warm-up on a treadmill or stationary bicycle, 5 min of stretching, 40 min of aerobic training, and 10 min of cool down and stretching. Subjects were permitted to select from a series of aerobic activities, for example, cycling on a stationary ergometer, running on a treadmill, climbing on a Stairmaster or using an elliptical trainer. Subjects exercised at a heart rate equivalent to 65–80% of their maximum heart rate obtained during CPET. Over the first 3 weeks, subjects gradually worked up to exercising at this intensity for 30 of the 40 min that they spent training on aerobic machines at each session. During weeks 4–12, they maintained this target heart rate for 30 min during every exercise session.

During training sessions, subjects wore a Polar HR monitor to record the HR during the entire session. After completing each training session, HR data were downloaded to a computer workstation. Study research assistants monitored these HR data to ensure that participants were exercising in their target zone. Participants assigned to the wait list condition maintained their usual lifestyle. After 12 weeks of training or wait list, subjects return for a second MRI scan, neuropsychology testing session and CPET. Participants assigned to the exercise

conditions received a 4-month membership in a fitness facility. To maintain a high degree of exercise adherence, participants who attended 90% of their scheduled 48 exercise sessions were awarded a bonus of two additional months of fitness center membership. Participants assigned to the wait list condition received 5 months of fitness center membership after completing their second testing session. All participants received $400 for the testing sessions.

Assessment of aerobic fitness. Maximum aerobic fitness (VO_2 max) was assessed by a 30-W (W), every 2 min, graded exercise test on an Ergoline 800S cycle ergometer (SensorMedics), until VO_2 max criteria (respiratory quotient _ 1.1, increases in ventilation without concomitant increases in VO_2, achievement of maximum age predicted heart rate, and/or volitional fatigue) were reached. The highest VO_2 value attained was considered VO_2 max[40]. Minute ventilation was measured by a pneumotachometer connected to a FLO-1 volume transducer module (PHYSIO-DYNE Instrument). Ventilatory gas analysis was assessed using paramagnetic O_2 and infrared CO_2 analyzers connected to a computerized system (MAX-1, Physiodyne Instrument). All systems were calibrated against known medical grade gases.

Statistical analysis. Prior to testing for group differences, scores on the ModBent task and VO_2 max at both baseline and follow-up were screened for outliers such that any values larger than 3 s.d. away from the mean were deleted from further analyses. One individual assigned to the high flavanol and exercise group had a ModBent value 3.55 and 4.06 s.d. above the mean at baseline and follow-up, respectively, and was deleted from further analysis concerning the ModBent. No other outliers were detected for ModBent nor for VO_2 max. The Shapiro-Wilk test was used to assess the normality distributional assumption for post treatment values of ModBent (after removing the outlier) and VO_2 max and found not to be rejected indicating the data did not violate normality assumptions.

From the $n = 41$ individuals randomized, there were $n = 35$ with both baseline and follow-up measures for ModBent (after taking out the one outlier, one individual was missing baseline and four individuals were missing follow-up), and $n = 35$ with complete data for VO_2 max (six individuals were missing follow-up). Complete case analysis of covariance (ANCOVA) was used to test randomized group differences in mean outcomes (ModBent and VO_2 max) by regressing post treatment measures on continuous baseline measures, dichotomous indicators for flavanol and exercise groups, and the interaction of flavanol and exercise indicators. Two-sided t tests for coefficients from ANCOVA were used to determine statistical significance at the alpha = 0.05 level. When the flavanol by exercise interaction was found not to be significant ($P > 0.05$), an additional ANCOVA was fit without the interaction to test the overall main effects of flavanol and exercise. Statistically significant main effects are presented by comparing adjusted mean outcomes post treatment for each group including standard error bars. ANCOVA-adjusted mean outcomes represent the expected mean post treatment for a person with a mean value of the measures at baseline. A similar ANCOVA model was used to study the effect of dietary intervention on the delayed retention score to establish the specificity of the findings.

As a robustness check for missing data (that is drop outs), ANCOVAs were repeated using 50 multiply imputed data sets (using Proc MI in SAS)[41] where data were imputed for individuals with a missing baseline or follow-up measure. Specifically, 50 data sets were generated with all missing data randomly imputed under a multivariate normality and missing at random assumption. ANCOVAs with and without interactions were again performed across all imputed data sets and results were combined and summarized using Proc MIANALYZE in SAS, which incorporates uncertainty due to the missing values. Through the generation of several imputed data sets and the subsequent reanalysis of each one of them, the uncertainty associated with the missing observations is accounted for by incorporating the error that arises from the fact that the regression coefficients differ across all of the different imputed data sets. Simulation studies routinely show decreased bias and improved efficiency using multiple imputation as compared to other techniques for handling missing data[42]. The findings for treatment effects on ModBent and VO_2 max were found to be qualitatively the same in terms of statistical significance using the multiple imputation technique.

A **Supplementary Methods Checklist** is available.

38. Reuter, M., Rosas, H.D. & Fischl, B. Highly accurate inverse consistent registration: a robust approach. *Neuroimage* **53**, 1181–1196 (2010).

39. Frangi, A., Niessen, W., Vincken, K. & Viergever, M. Multiscale vessel enhancement filtering. *Med. Image Comput. Comput. Assist. Interv.* **1496**, 130–137 (1998).

40. Buchfuhrer, M.J. *et al.* Optimizing the exercise protocol for cardiopulmonary assessment. *J. Appl. Physiol.* **55**, 1558–1564 (1983).

41. Yuan, Y.C. Multiple imputation for missing data: concepts and development (Version 9.0). *SAS Support* http://support.sas.com/rnd/app/stat/papers/multipleimputation.pdf (2012).

42. Little, R.J.A. & Rubin, D.B. *Statistical Analysis with Missing Data* (Wiley, Hoboken, N.J., 2002).