**Increased Neural Differentiation after a Single Session of Aerobic Exercise in Older Adults.**

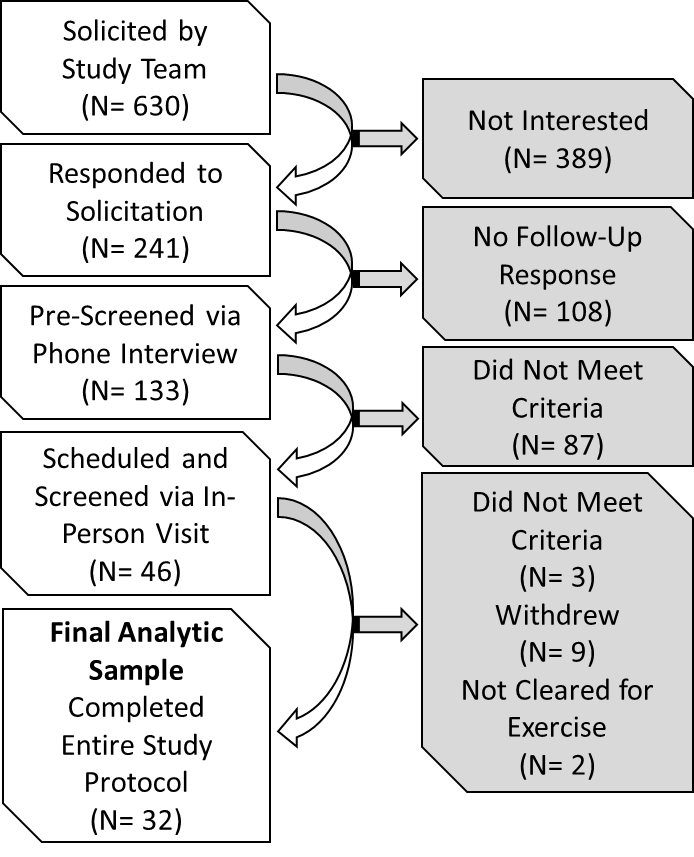
Jeremy Purcell1,3, Junyeon Won2, Daniel Callow1,3, Robert Wiley5, Lauren Weiss3, Alfonso Alfini6, and J. Carson Smith1,3

1. Department of Kinesiology, University of Maryland, College Park, MD, 20740, USA
2. Institute for Exercise and Environmental Medicine, Texas Health Presbyterian Dallas, Dallas, TX 75231, USA
3. Program in Neuroscience and Cognitive Science, University of Maryland, College Park, MD, 20740, USA
4. Maryland Neuroimaging Center, University of Maryland, College Park, MD, 20740, USA
5. Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC, 27402, USA
6. National Center on Sleep Disorders Research, Division of Lung Diseases, National Heart, Lung, and Blood Institute, Bethesda, MD

**\*corresponding authors:**

Jeremy Purcell: Email**:**  [jpurcel8@umd.edu](mailto:jpurcel8@umd.edu)

J. Carson Smith: [carson@umd.edu](mailto:carson@umd.edu)



**Figure S1**: A. Study recruitment and enrollment summary (2016 – 2018).

**Text S1 Materials and Methods: fMRI Task Description**

Task run paradigms were administered electronically using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). These included the following four paradigms, the durations of which are summarized in Table 2. (1) Flanker task: An event-related version of the Eriksen Flanker task (Colcombe et al., 2004; Eriksen and Eriksen, 1974). Each trial began with a 1500–9500 ms central fixation screen followed by a dashed line cue (500 ms) to indicate the stimulus was about to appear. The stimulus was a centrally presented row of 5 arrow stimuli, presented for 500 ms each and followed by a 1500 ms response window. The stimulus arrows could all be facing the same direction (congruent) or the central arrow could be facing the opposite direction as the other four (incongruent). There were a total of 80 trials including 40 congruent and 40 incongruent trials; an optimized schedule of trial presentation was generated using Optseq2. Participants were instructed to respond to the direction of a central arrow with their right hand using their index or middle finger. (2) International Affective Pictures (IAPS) task: A blocked-design paradigm that involved viewing  pictures from the International Affective Picture System (IAPS) (Lang et al., 2008). A total of 120 pictures from the IAPS were presented, including 40 pictures each of neutral, pleasant and unpleasant valence categories. Each picture was presented for 2 seconds, and organized into blocks of 12 images. Stimulus blocks were randomly order with a 22 second rest period between them during which a centrally located fixation cross was presented. Participants were instructed to press a button using the index finger of their right hand to indicate indoor images and another button using the middle finger of their right hand for outdoor images. (3) Resting State: Participants were instructed to keep their eyes open, look at a central fixation cross, and try not to fall asleep for the duration of the scan. (4) Famous Names Task (FNT): An event-related design paradigm that involved viewing famous and non-famous names, adapted from studies of semantic memory (Smith et al., 2013; Sugarman et al., 2012). A total of 60 names were visually presented: 30 famous (e.g., Frank Sinatra) and 30 non-famous (Douville et al., 2005) Stimuli were presented for 4 seconds each with randomly interspersed 4-second centrally located fixation rest periods that took up 1/3 of the timepoints. Participants were instructed to press a button using the index finger of their right hand for Famous names and another button using the middle finger of their right hand for Non-Famous names.

To minimize practice effects across sessions due to the same condition or stimulus sequences, each participant saw a unique version of each functional task at each visit (e.g, the condition orders of the Flanker task were not the same across visits).

**Text S2 Results: In Scanner Task Performance**: In scanner task performance for the Flanker Task and the Famous Names Task are reported in previous studies (Won et al., 2019b, 2019a), and is not discussed further here because the fMRI task performance data were accounted for as regressor covariates, and were thus ignored in this analysis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table S1. Summary of fMRI Run Durations.** | | | | |
| Functional Run # | Task | Final Analyzed Data | | Number (%) of Participants with runs at both Rest and Exercise Sessions |
| # of Time Points | Duration (Min:Sec) |
| 1 | Flanker | 284 | 8:04 | 32 (100%) |
| 2 | IAPS | 242 | 7:52 | 29 (91%) |
| 3 | Resting | 236 | 0:00 | 32 (100%) |
| 4 | FNT | 171 | 5:42 | 26 (81%) |
| **Pre Motion Censor** | |  |  |  |
| Max Total Time | | 933 | 31:06 | 24 (75%) |
| Min Total Time | | 521 | 17:22 | 1 (3%) |
| **Mean Total Time** | | **887** | **29:34** |  |
| **Post Motion Censor** | |  |  |  |
| Max Total Rest Time | | 933 | 31:06 |  |
| Min Total Rest Time | | 465 | 15:30 |  |
| Mean Total Rest Time | | 847.75 | 28:15 |  |
| Max Total Exercise Time | | 933 | 28:15 |  |
| Min Total Exercise Time | | 520 | 17:20 |  |
| Mean Total Exercise Time | | 853.25 | 28:26 |  |
| **Paired t-test of Total Time: Rest minus Exercise** | | **t= -1.1, p = 0.299** | |  |
|  |
| Notes: Final Analyzed Data = data after the initial 4 timepoints were removed and timepoints with > .5 scan to scan motion were censored. Pre and post motion censored timepoint amount are listed above. Total Time = the total fMRI data for participants *after* the available runs were concatenated; There were equal numbers of runs at both the Rest and Exercise sessions for all participants (i.e., if one participant had a run at the Rest session missing, that equivalent run was ignored at the Exercise session for that participant). IAPS, International Affective Picture System; FNT, Famous Name Task. Some runs were excluded due to technical issues at the time of acquisition or due to excessive motion. Most (75%) of the participants had the full 933 timepoints (31:06 minutes of data) pre-censoring. After censoring participants had an average of ~28 minutes of data for either the Rest of Exercise conditions. There was no significant difference between the amount of data post-censoring between the Rest and Exercise conditions. | | | | |
|
|
|
|

**Text S3 Materials and Methods: Intra-Class Correlation (ICC) Analysis**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S2** Experimental Manipulation Check (N = 32) | | | | | | | |
|  | | Rest | Exercise | Exercise - Rest | | | |
| mean (SEM) | mean (SEM) | Exercise minus Rest | Cohen's d | t-stat | p |
|  | Start Session HR | 70.3 (1.7) | 76.3 (1.8) |  |  |  |  |
|  | End Session HR | 66.6 (1.5) | 137.3 (3.4) |  |  |  |  |
|  | **HR Change** | **-3.7 (1)** | **+61 (3.3)** | **65 (58.1 - 71.5)** | **3.5** | **19.8** | **<.00001** |
|  | Start Session RPE | 6.2 (0.1) | 6.3 (0.1) |  |  |  |  |
|  | End Session RPE | 6.1 (0.1) | 14.6 (0.2) |  |  |  |  |
|  | **RPE Change** | **-0.1 (0.1)** | **+8 (0.3)** | **8 (7.9 - 9)** | **5.6** | **31.8** | **<.00001** |
| Notes: SEM = Standard Error of the Mean; Heart Rate Change (BPM) = the change in heart rate in beats per minute from the beginning to the end of either the 30 minute Rest or the Exercise period; RPE change = the change in the Rating of Perceived Exertion from the beginning to the end of either the 30 minute Rest or Exercise period. | | | | | | | |

Whole brain and regional intra-class correlations (ICCs) were calculated (Chen et al., 2018). Given there were four functional runs, we opted to look at split half reliability by concatenating pairs of functional runs, and then computing the ICC’s based on whether two pairs of concatenated runs reproduced similar values. For instance, numbered in Table 2, we concatenated the Flanker (1) and IAPS (2), and calculated the ICC values relative to the concatenated Resting State (3) and FNT (4) runs. This was done for every unique pair-wise combination of concatenated runs (i.e., runs 1+2 to 3+4, runs 1+3 to 2+4, and runs 1+4 to 2+3, where the + symbol refers to concatenation of the labeled runs). This approach gives 3 separate whole brain ICC maps that correspond to each unique pairwise combination. The final summary ICC value is calculated as the mean of these 3 maps. In order to account for differences in the number of time points in the different run combinations, the whole brain (i.e., gray matter) General-Hreg and the total time point number were included as a covariates in the 3dICC analysis. As with the General-Hreg analysis above, the whole brain General-Hreg was also included as a covariate. A group-level ICC(2,1) was calculated using AFNI’s (3dICC) as discussed here (Chen et al., 2018). This ICC provides a measure of absolute agreement of the different functional runs. That is, higher ICC’s correspond to when the different functional runs produce the same/similar *magnitude* of value - as opposed to the more relative ICC(3,1) which is more analogous to a Pearson correlation and thus independent of absolute magnitude similarities (Elliott et al., 2020).

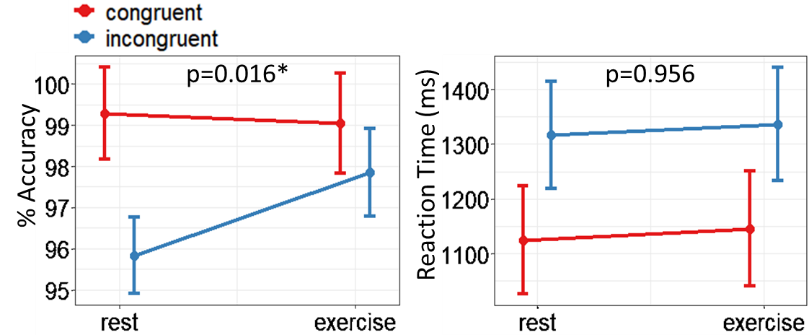
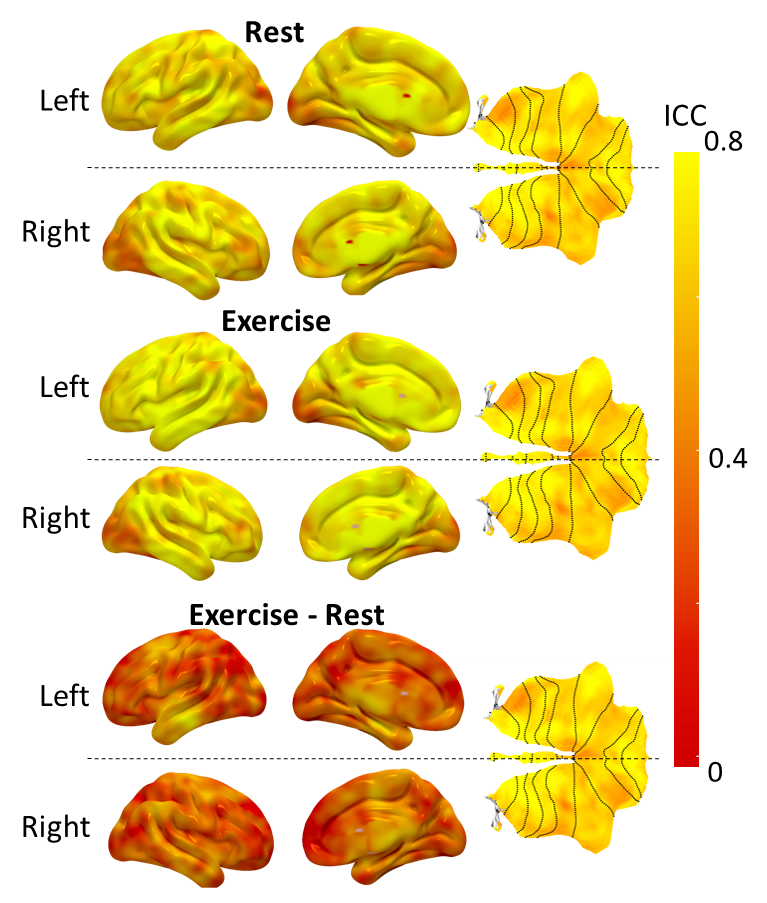


Figure S2. Stroop Behavior Results. As reported in Table 3, there was a significant interaction effect between condition (rest vs exercise) and Stroop task trial type (congruent vs incongruent trials) for % Accuracy, but not for Reaction time (noted by the crossed lines in the left plot and parallel lines in the right plot).

****

FigureS3. General-Hreg group-level split-half reliability Intra-Class Correlation (ICC) maps. The color scale depicts a maximum of 0.8 (strong reliability) to a minimum of 0 (no reliability). For reference, we use the following scale: [0-0.4] = poor; [0.4-0.6] = fair; [0.6-0.75] = good; [0.75-1] = strong. As also reported in Table 4, although the Exercise – Rest condition maps had reliability that mostly rated in the poor to fair range, there were regions with good reliability (yellow regions). The within condition maps on the other hand, (top Rest and middle Exercise) had mostly good to strong reliability.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table S3: Fixed effects outputs from the 6 separate mixed effects models used to quantify the interaction between age and the Experimental conditions (Rest and Exercise). Models use mean region of interest values, unlike those reported in Table and Figure 4.** | | | | |
| Regions of Interest | Fixed Effects of Interest | t | p (Sig.) | R ² Full Model |
| Cerebellum Crus I | Age | 0.062 | 0.9509 | 0.9200 |
| Rest\_Exercise:Age | 1.637 | 0.1120 |  |
| **Rest\_Exercise:Age:Region:Hemisphere** | **-1.199** | **0.0283 \*** |  |
| Cerebellum Crus II | Age | -1.057 | 0.2986 | 0.9119 |
| Rest\_Exercise:Age | 1.508 | 0.1423 |  |
| Rest\_Exercise:Age:Region:Hemisphere | -1.695 | 0.1004 |  |
| Cerebellum Lobule VI | Age | -0.293 | 0.7713 | 0.7134 |
| Rest\_Exercise:Age | 1.454 | 0.1568 |  |
| Rest\_Exercise:Age:Region:Hemisphere | -0.151 | 0.2320 |  |
| Vermis | Age | -0.363 | 0.7188 | 0.9136 |
| Rest\_Exercise:Age | 1.493 | 0.1461 |  |
| Rest\_Exercise:Age:Region:Hemisphere | 0.089 | 0.9296 |  |
| Parahippocampal Gyrus /Entorhinal Cortex | Age | 0.748 | 0.4610 | 0.8051 |
| Rest\_Exercise:Age | 0.632 | 0.5320 |  |
| Rest\_Exercise:Age:Region:Hemisphere | 0.541 | 0.5920 |  |
| Hippocampus | Age | -0.522 | 0.6057 | 0.9262 |
| Rest\_Exercise:Age | 1.932 | 0.0629. |  |
| Rest\_Exercise:Age:Region:Hemisphere | 0.644 | 0.5246 |  |
| Notes: Tabulated results from the six separate post hoc region of interest mixed effect model analyses; see Figure 4 for plotted results. The model regions are listed on the left. For each left and right region the General-Hreg values were the mean values across all voxels in each region for each subject. These General-Hreg values were set as the dependent variables in each of the 6 models. This resulted in six mixed effects models that quantified the fixed effects of age, condition, the two hemispheres and all interactions between them and also a full set random effect slopes and intercepts. Effects of Age and Rest\_Exercise Conditions interactions are of interest and reported here; the main effect of Rest\_Exercise is not reported due to the selection process explicitly selecting voxels that are different across these conditions. Significance: \* significant at alpha = 0.05; · approaching significance at p<0.6. Although only the cerebellar Crus 1 interactions are significant, the patterns observed in the Crus 2 and PHG/EC match those of the more refined approach reported in Table and Figure 4. This provides some qualitative support for the original hypothesis that those voxels with the greatest Exercise versus Rest difference in differentiation, would be the ones that interact most with age. | | | | |

**SI References**

Chen, G., Taylor, P.A., Haller, S.P., Kircanski, K., Stoddard, J., Pine, D.S., Leibenluft, E., Brotman, M.A., Cox, R.W., 2018. Intraclass correlation: Improved modeling approaches and applications for neuroimaging. Hum Brain Mapp 39, 1187–1206. https://doi.org/10.1002/hbm.23909

Colcombe, S.J., Kramer, A.F., Erickson, K.I., Scalf, P., McAuley, E., Cohen, N.J., Webb, A., Jerome, G.J., Marquez, D.X., Elavsky, S., 2004. Cardiovascular fitness, cortical plasticity, and aging. PNAS 101, 3316–3321. https://doi.org/10.1073/pnas.0400266101

Douville, K., Woodard, J.L., Seidenberg, M., Miller, S.K., Leveroni, C.L., Nielson, K.A., Franczak, M., Antuono, P., Rao, S.M., 2005. Medial temporal lobe activity for recognition of recent and remote famous names: an event-related fMRI study. Neuropsychologia 43, 693–703.

Elliott, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Sison, M.L., Moffitt, T.E., Caspi, A., Hariri, A.R., 2020. What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. Psychol Sci 31, 792–806. https://doi.org/10.1177/0956797620916786

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. Perception & psychophysics 16, 143–149.

Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2008. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. (No. A-8). University of Floriday, Gainesville, FL.

Smith, J.C., Nielson, K.A., Antuono, P., Lyons, J.-A., Hanson, R.J., Butts, A.M., Hantke, N.C., Verber, M.D., 2013. Semantic memory functional MRI and cognitive function after exercise intervention in mild cognitive impairment. Journal of Alzheimer’s Disease 37, 197–215.

Sugarman, M.A., Woodard, J.L., Nielson, K.A., Seidenberg, M., Smith, J.C., Durgerian, S., Rao, S.M., 2012. Functional magnetic resonance imaging of semantic memory as a presymptomatic biomarker of Alzheimer’s disease risk. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1822, 442–456.

Won, J., Alfini, A.J., Weiss, L.R., Callow, D.D., Smith, J.C., 2019a. Brain activation during executive control after acute exercise in older adults. Int J Psychophysiol 146, 240–248. https://doi.org/10.1016/j.ijpsycho.2019.10.002

Won, J., Alfini, A.J., Weiss, L.R., Michelson, C.S., Callow, D.D., Ranadive, S.M., Gentili, R.J., Smith, J.C., 2019b. Semantic Memory Activation After Acute Exercise in Healthy Older Adults. Journal of the International Neuropsychological Society 1–12.