Physical Exercise Performed Four Hours after Learning Improves Memory Retention and Increases Hippocampal Pattern Similarity during Retrieval

Physical exercise performed four hours after learning improves the retention of associative memories and increases hippocampal pattern similarity during retrieval

Eelco V. van Dongen[12], Ingrid H.P. Kersten[2], Isabella C. Wagner[12], Richard G. M. Morris[3], Guillén Fernández[12]

[1]Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, PO Box 9101, 6500 HB, Nijmegen, The Netherlands.

[2]Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, PO Box 9101, 6500 HB, Nijmegen, The Netherlands

[3]Centre for Cognitive and Neural Systems, The University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom.

Contact: Guillén Fernández, guillen.fernandez@donders.ru.nl

Highlights

- Performing aerobic exercise four hours after learning improved associative memory
- Exercise at this time also increased hippocampal pattern similarity during retrieval
- Exercise performed immediately after learning had no effect on memory retention
- Exercise could have potential as a memory intervention in educational settings

eTOC blurb

Van Dongen and colleagues show that acute aerobic exercise performed four hours, but not directly, after learning improves long-term memory in humans. Such exercise was also associated with higher consistency in hippocampal activation during memory recall. These findings suggest that correctly timed exercise holds promise as a memory intervention.

Summary

Persistent long-term memory depends on successful stabilization and integration of new memories after initial encoding [1, 2]. This process of consolidation is thought to require several neuromodulatory factors such as dopamine, noradrenaline and brain-derived neurotrophic factor [3-7]. Without the release of such factors around the time of encoding, memories will decay rapidly [3, 5, 6, 8]. Recent studies have shown that physical exercise acutely stimulates the release of several consolidation-promoting factors in humans [9-14], raising the question whether physical exercise can be used to improve memory retention [15-17]. Here, we used a single session of physical exercise after learning to exogenously boost memory consolidation and thus long-term memory. Three groups of randomly-assigned participants first encoded a set of picture-location associations. Afterwards, one group performed exercise immediately, one four hours later, and the third did not perform any exercise. Participants otherwise underwent exactly the same procedures to control for potential experimental confounds. Forty-eight hours later, participants returned for a cued recall test in a magnetic resonance scanner. With this design we could investigate the impact of acute exercise on memory consolidation and retrieval-related neural processing. We found that performing exercise four hours but not immediately after encoding improved the retention of picture-location associations compared to the no-exercise control group. Moreover, performing exercise after a delay was associated with increased hippocampal pattern similarity for correct responses during delayed retrieval. Our results suggest that appropriately timed physical exercise can improve long-term memory and highlight the potential of exercise as an intervention in educational and clinical settings.

Results

72 participants were randomly assigned to one of three age- and gender-matched groups; all learned 90 picture-location associations over a period of approximately 40 minutes (**Figure 1;** for visualization of the experimental trials, see **Figure S1;** for sample demographics, see **Table S1**). In each group, half the participants started at 9AM and half at 12PM to control for time-of-day effects. Following a baseline cued recall test (*Test 1*), participants in the Immediate Exercise group (IE) performed a 35 minute interval training on an ergometer at an intensity of up to 80% of their maximum heart rate (**See Supplemental Experimental Procedures** and **Figure S2**). IE participants subsequently moved to a separate quiet environment for a 3-hour delay period, where they watched nature documentaries, before returning to the exercise lab for a control session. This control session did not involve exercise but used the same context otherwise. For the Delayed Exercise group (DE), the protocol was identical, but with the order of the exercise and control session reversed; for the No Exercise group (NE), both sessions before and after the delay period were control sessions. Participants returned to the lab 48 hrs after initial encoding and performed a second cued recall test (*Test 2*) in the MR scanner. With this design, we could investigate whether post-learning physical exercise affected memory retention, if its effects were time-dependent, and whether our intervention influenced the neural substrate of memory retrieval as measured using fMRI.

Response to exercise

Our exercise intervention was successful in raising our participants' heart rates and subjective ratings of exercise intensity, in line with the intended schedule (see **Supplemental Experimental Procedures, Figure S2 and Figure S3**), indicating that participants experienced the expected physiological consequences of the interval training.

Behavioral performance

Memory performance on Test 1 was well above chance level for all experimental groups (one-sample T-test versus chance level (16.67%); percentage of correct responses ± standard error of the mean (SEM): μ_{NE}=79.1±3.8%, μ_{IE}=79.3±2.7%, μ_{DE}=85.2±2.4%; t_{NE}=16.56, t_{IE}=23.00, t_{DE}=28.46; all $p<0.001$, see also **Figure S4),** indicating that all groups successfully learned a substantial number of picture-location associations. Performance at Test 1 was not significantly different between groups **(**F_{Group}=1.48, p_{Group}=0.236) nor dependent on starting time or gender (F_{Time}=0.235, p_{Time}=0.630; F_{Gender}=1.74, p_{Gender}=0.192).

To find out if exercise influenced the consolidation of picture-location associations, and thus the retention of information learned on Day 1, we used memory retention as our primary memory measure. This measure was defined as Test 2 performance divided by Test 1 performance, and thus corrected for baseline performance differences between participants. Memory retention was significantly different between groups (μ_{NE}=0.800, μ_{IE}=0.795, μ_{DE}=0.866; F_{Group}=4.83, p_{Group}=0.011, see **Figure 2**). Post-hoc pair-wise comparisons showed that retention in the DE group was higher than in the IE and NE groups, with no difference between the latter two groups (DE-IE: $p_{TukeyHSD}$=0.031; DE-NE: $p_{TukeyHSD}$=0.045; IE-NE: $p_{TukeyHSD}$=0.986). Raw performance scores for Test 1 and Test 2 can be found in **Figure S4** and **Table S1**.

In addition, memory retention was higher in female than male participants (mean ± SEM: μ_{Female}=0.84±0.015, μ_{male}=0.78±0.021, F_{Gender}=6.32, p_{Gender}=0.019) but this effect did not interact with the observed retention difference between groups ($F_{GroupxGender}$=0.639, $p_{GroupxGender}$=0.531). Similarly, no main effect of participants' starting time nor interaction between group and starting time was observed (F_{Time}=3.21; p_{Time}=0.078; $F_{GroupxTime}$=2.31, $p_{GroupxTime}$=0.108). Memory retention was not correlated with participants' weekly exercise duration or frequency (Pearson's $r_{duration}$= 0.057, $p_{duration}$= 0.636; Pearson's $r_{frequency}$= 0.031, $p_{frequency}$=0.799, see also **Table S2**).

To investigate if exercise additionally modulated retrieval time and/or subjective measures of memory strength, we also analyzed the reaction times and confidence ratings during recall. Both reaction times and confidence ratings during Test 1 and Test 2 were not significantly different between groups (all p>0.05, for the raw values see **Table S1**). Across the sample, confidence was higher for correct versus incorrect responses at both Test 1 and Test 2 (Paired Samples T-tests, Test 1: $\mu_{correct}$=4.41, $\mu_{incorrect}$=2.20, t_{Test1}=22.95, p_{Test1}<0.001; Test 2: $\mu_{correct}$=4.18, $\mu_{incorrect}$=2.24, t_{Test2}=26.14, p_{Test2}<0.001). Similarly, reaction times were shorter for correct than incorrect responses (Paired Samples T-tests, Test 1: $\mu_{correct}$=2134ms, $\mu_{incorrect}$=2711ms, t_{Test1}=8.15, p_{Test1}<0.001; Test 2: $\mu_{correct}$=2646ms, $\mu_{incorrect}$=3012ms, t_{Test2}=6.23, p_{Test2}<0.001). Reaction times could not be directly compared between Test 1 and Test 2 since the MR scanner environment of Test 2 was substantially different from Test 1's experimental lab setting. However, confidence levels for correct responses were lower at Test 2 versus Test 1 (Paired Samples T-tests, $\mu_{Test1correct}$=4.41, $\mu_{Test2correct}$=4.18, t=7.85, p<0.001).

Together these results suggest that performing physical exercise four hours after encoding promoted the retention of associative memory, without differentially affecting participants' reaction times or confidence levels.

Functional neuroimaging results

Using functional imaging, we found that BOLD activation was increased in a wide range of brain regions during correct recall relative to a fixation baseline (see **Figure 3A**). Moreover, our analyses showed a significant difference in BOLD activation during correct and incorrect responses in bilateral hippocampus, the striatum, and prefrontal, occipital and parietal areas (see **Figure 3B and Table S3**). However, we observed no significant effect of experimental group for either contrast (for analytic details, see **Supplemental Experimental Procedures**).

Based on previous research demonstrating the relevance of consistent neural processing (i.e. pattern similarity) for memory retention, and its reported utility in investigating differences in neural representations between experimental conditions and groups, we then conducted a hippocampal pattern similarity analysis [18-22]. We speculated that exercise-related physiological effects (e.g. dopamine and/or noradrenaline) could have neuromodulatory effects and thus alter the neural representations of recently encoded memories [23, 24]. In this way, exercise might produce differences in the neural response patterns observed during recall instead of differences in regional BOLD amplitude.

Using regions of interest in left and right hippocampus (**Supplemental Experimental Procedures**), we found that hippocampal pattern similarity was significantly different between groups during correct trials (**Figure 4**). A three-way Repeated Measures ANOVA with the factors Group, Correctness and Hemisphere (left/right) indicated a main effect of Group and Correctness, and a significant GroupxCorrectness interaction (F_{Group}=5.35, p_{Group}=0.007; $F_{Correct}$=3.97, $p_{Correct}$=0.050; $F_{GroupxCorrect}$=5.43, $p_{GroupxCorrect}$=0.007, all other effects and interactions p>0.05). Post-hoc pair-wise comparisons indicated that hippocampal pattern similarity during correct (vs. incorrect) trials was higher for the DE group than both IE and NE groups and not different between IE and NE participants (DE-IE: p_{Sidak}=0.012; DE-NE:p_{Sidak}=0.003; IE-NE: p_{Sidak}=0.967). Pattern analyses in other brain regions showed that significant group differences were limited to the hippocampus only (**Figure S5**). Pearson's correlations between hippocampal pattern similarity and memory retention showed that across the sample, higher hippocampal pattern similarity was weakly but significantly associated with better memory retention (Pearson's r=0.287; p=0.015).

Discussion

Together, these results indicate that performing physical exercise after learning can improve the retention of associative memories, and modulates the consistency of hippocampal activation patterns during retrieval.

Considering that the exercise intervention took place after learning, delayed exercise most likely affected memory retention through an impact on memory consolidation. As such, it seems likely that one or more of the physiological consequences of aerobic exercise facilitated consolidation. Although we did not measure this directly in our study, previous research suggests that exercise triggers the release of BDNF, plasticity-related products (PRPs), noradrenaline and dopamine, amongst other substances that promote neural plasticity. Such factors are critical for the consolidation of synaptic potentiation, as proposed in the Synaptic Tagging and Capture (STC) hypothesis [3, 8] and are also important for later stages of memory consolidation [25, 26]. One possibility is that the release of PRPs at a time where naturally lower levels of PRPs would be available (i.e. several hours after learning) could have mediated the facilitation of memory retention in our study. Alternatively, exercise-dependent release of dopamine and noradrenaline could have facilitated consolidation similar to previously described effects of novelty and arousal [4, 5, 7, 14, 24, 27, 28].

We found no evidence for any effect of physical exercise immediately after learning, suggesting that the physiological response to exercise did not benefit memory consolidation at this stage. This finding is not predicted by consolidation theories such as the STC hypothesis. One explanation for this result could be that the neural context at this time was already optimal following initial learning and recall, and could not be further improved through an additional influx of consolidation-promoting factors. Indeed, the good performance in all three groups at Test 1 suggests that the study procedure itself enabled high levels of recall initially. Alternatively, the time course of synaptic consolidation might be different in humans compared to animals. Our experimental design does not allow us to directly investigate these explanations, however, so they should be considered as speculative until supported by other studies. And even though our results do not seem to directly support the simplest STC prediction, the lack of

current knowledge about the time course and mechanisms of STC in humans warrants caution in interpreting our findings this way.

We used a declarative memory task. However, previous studies using procedural tasks have provided evidence that in those settings the close proximity of exercise maximizes its effects on memory [29-31](for an exception, see [32]). For this reason, it seems that the effect of exercise on memory is not only modulated by timing, but also by the type of memory investigated. Several studies using declarative tasks have also shown effects of exercise conducted immediately after acquisition, suggesting that experimental parameters such as task, type of learning, type of stimulus material and subject population could be important for the net effect of exercise on memory retention [12, 14, 33, 34]. It remains a challenge for future research to determine the specific parameters that modulate the impact of exercise on memory.

Univariate analyses of BOLD fMRI activity during Test 2 showed no effects of exercise. Although contrasts between correct and incorrect responses did provide evidence that brain regions associated with memory recall were recruited during our task, they did not distinguish between experimental groups. This finding suggests that cued recall after delayed exercise does not involve different brain circuits compared to no or immediate exercise. Moreover, delayed exercise was apparently not related to an overall modulation of BOLD signal amplitude in the correct versus incorrect responses contrast. Instead, our pattern similarity analysis points to delayed exercise might having caused a qualitative change in the activation patterns associated with correct recall of the picture-location associations.

As far as we know, between-group differences in hippocampal pattern similarity during retrieval in regards to acute exercise have not been reported previously. They could represent differential efficiency or coherence during memory retrieval and might relate to differences in memory strength [19]. The effect of exercise on pattern similarity could alternatively be interpreted as an increase in the signal to noise ratio, which is intriguing considering the neuromodulatory roles of dopamine and noradrenalin and their known association with exercise [9, 10, 13, 23, 24]. Interestingly, although pattern similarity was, across our sample, higher for correct compared to incorrect responses in hippocampus, striatum and medial prefrontal cortex, it was specifically increased for the DE group in hippocampus only. This finding could imply that the hippocampus is particularly sensitive to the acute effects of exercise, and thus important in mediating its cognitive benefits. The correlation between hippocampal pattern similarity and memory retention seems in line with this prediction, but more research is needed to substantiate this claim.

The results of the current study should be seen in the context of some limitations. First, we did not measure any peripheral or central measures of BDNF, dopamine, noradrenalin or other factors released during aerobic exercise, and thus cannot conclude with certainty that such factors mediated the behavioral and neural effects of our exercise intervention. Future studies should ideally include such measures or specific experimental manipulations to gain more insight into the molecular mechanisms of exercise-related memory improvement.

Second, based on the current data, we cannot yet determine the exact time window in which delayed exercise is effective in promoting memory retention. Future studies should include experimental groups that perform exercise at time points beyond 4 hours after learning to better delineate the critical time period for these effects.

Third, it is not possible to determine whether the observed effects of exercise require sleep and/or prolonged consolidation, or could already have been observed during or shortly after exercise. Future research should further investigate to what degree sleep and time contribute to the mnemonic effects of acute aerobic exercise. Related to this, it is not yet clear how exercise affects memory retention beyond 48 hours after learning. Follow-up research should include longer retention periods to see whether or not positive effects of delayed exercise persist beyond 2 days.

Finally, we would like to stress that much is yet unknown about the molecular mechanisms of consolidation in humans. Not only is the timing of early consolidation processes as yet poorly understood, we do not know whether the molecular factors critical in animal models play similar roles in humans. In addition, by necessity, human studies generally use correlational analyses and peripheral measures of the physiological effects of exercise when investigating the mnemonic impact of exercise. It remains unclear how serum levels of e.g. BDNF, dopamine and noradrenaline relate to local changes in consolidation factors in the neural circuits important for long-term memory. As such, our speculations on possible mechanistic explanations of our findings should be interpreted with caution.

Regardless of these limitations, our results provide initial evidence that properly timed physical exercise can alter mnemonic processes at delayed retrieval and improve memory retention over a period of at least 48 hours. This finding is in line with previous studies reporting beneficial effects of physical exercise (for a review, see[15]) and highlights its potential as a memory intervention in humans. The economic, healthy and practical nature of exercise makes it ideal for interventions in educational and clinical settings. Our experiment thus serves as a proof of principle study that could inspire future applications of exercise to boost long-term memory in various populations.

Contributions

E.V.D., R.G.M.M. and G.F. designed the experiment. E.V.D. and I.H.P.K. acquired the data. E.V.D. and I.H.P.K. analyzed the data. I.C.W. created an analysis toolbox for the pattern similarity analyses and assisted with this part of the analysis. E.V.D., I.H.P.K., I.C.W., R.G.M.M. and G.F. wrote the paper.

Acknowledgements

R.G.M.M. and G.F. were supported by a grant from the European Research Council, ERC AdG2010 – 268800 – NEUROSCHEMA. The authors would like to thank Maria Hopman for helpful advice during the setup of this study, and Arie van Dijk for clinical assistance during the ECG screening.

References

1. Dudai, Y. (2012). The Restless Engram: Consolidations Never End. Annu Rev Neurosci. *35,* 227-247
2. Squire, L.R., Genzel, L., Wixted, J.T., and Morris, R.G. (2015). Memory Consolidation. Cold Spring Harb Perspect Biol *7*.
3. Redondo, R.L., and Morris, R.G. (2011). Making memories last: the synaptic tagging and capture hypothesis. Nat Rev Neurosci *12*, 17-30.
4. Moncada, D., Ballarini, F., Martinez, M.C., Frey, J.U., and Viola, H. (2011). Identification of transmitter systems and learning tag molecules involved in behavioral tagging during memory formation. Proc Natl Acad Sci U S A *108*, 12931-12936.
5. Moncada, D., and Viola, H. (2007). Induction of long-term memory by exposure to novelty requires protein synthesis: evidence for a behavioral tagging. J Neurosci *27*, 7476-7481.
6. Wang, S.H., Redondo, R.L., and Morris, R.G. (2010). Relevance of synaptic tagging and capture to the persistence of long-term potentiation and everyday spatial memory. Proc Natl Acad Sci U S A *107*, 19537-19542.
7. McGaugh, J.L. (2006). Make mild moments memorable: add a little arousal. Trends Cogn Sci *10*, 345-347.
8. Frey, U., and Morris, R.G. (1997). Synaptic tagging and long-term potentiation. Nature *385*, 533-536.
9. Winter, B., Breitenstein, C., Mooren, F.C., Voelker, K., Fobker, M., Lechtermann, A., Krueger, K., Fromme, A., Korsukewitz, C., Floel, A., et al. (2007). High impact running improves learning. Neurobiol Learn Mem *87*, 597-609.
10. Skriver, K., Roig, M., Lundbye-Jensen, J., Pingel, J., Helge, J.W., Kiens, B., and Nielsen, J.B. (2014). Acute exercise improves motor memory: exploring potential biomarkers. Neurobiol Learn Mem *116*, 46-58.
11. Knaepen, K., Goekint, M., Heyman, E., and Meeusen, R. (2010). Neuroplasticity — Exercise-Induced Response of Peripheral Brain-Derived Neurotrophic Factor. Sports Med *40*, 765-801.
12. Griffin, E.W., Mullally, S., Foley, C., Warmington, S.A., O'Mara, S.M., and Kelly, A.M. (2011). Aerobic exercise improves hippocampal function and increases BDNF in the serum of young adult males. Physiol Behav *104*, 934-941.
13. Rostrup, M., Westheim, A., Refsum, H.E., Holme, I., and Eide, I. (1998). Arterial and venous plasma catecholamines during submaximal steady-state exercise. Clin Physiol *18*, 109-115.
14. Segal, S.K., Cotman, C.W., and Cahill, L.F. (2012). Exercise-induced noradrenergic activation enhances memory consolidation in both normal aging and patients with amnestic mild cognitive impairment. J Alzheimers Dis *32*, 1011-1018.
15. Roig, M., Nordbrandt, S., Geertsen, S.S., and Nielsen, J.B. (2013). The effects of cardiovascular exercise on human memory: a review with meta-analysis. Neurosci Biobehav Rev *37*, 1645-1666.
16. Lojovich, J.M. (2010). The relationship between aerobic exercise and cognition: is movement medicinal? J Head Trauma Rehabil *25*, 184-192.
17. McMorris, T. (2009). Exercise and Cognitive Function: A Neuroendocrinological Explanation. In Exercise and Cognitive Function. (John Wiley & Sons, Ltd), pp. 41-68.
18. Hsieh, L.T., Gruber, M.J., Jenkins, L.J., and Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. Neuron *81*, 1165-1178.
19. Davis, T., Xue, G., Love, B.C., Preston, A.R., and Poldrack, R.A. (2014). Global neural pattern similarity as a common basis for categorization and recognition memory. J Neurosci *34*, 7472-7484.

20. Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J.A., and Poldrack, R.A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. Science *330*, 97-101.
21. Jimura, K., and Poldrack, R.A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia *50*, 544-552.
22. Xue, G., Dong, Q., Chen, C., Lu, Z.L., Mumford, J.A., and Poldrack, R.A. (2013). Complementary role of frontoparietal activity and cortical pattern similarity in successful episodic memory encoding. Cereb Cortex *23*, 1562-1571.
23. Hurley, L.M., Devilbiss, D.M., and Waterhouse, B.D. (2004). A matter of focus: monoaminergic modulation of stimulus coding in mammalian sensory networks. Curr Opin Neurobiol *14*, 488-495.
24. Mather, M., Clewett, D., Sakaki, M., and Harley, C.W. (2015). Norepinephrine ignites local hot spots of neuronal excitation: How arousal amplifies selectivity in perception and memory. Behav Brain Sci, 1-100.
25. Bekinschtein, P., Cammarota, M., and Medina, J.H. (2014). BDNF and memory processing. Neuropharmacology *76, Part C*, 677-683.
26. Katche, C., Cammarota, M., and Medina, J.H. (2013). Molecular signatures and mechanisms of long-lasting memory consolidation and storage. Neurobiol Learn Mem *106*, 40-47.
27. Shohamy, D., and Adcock, R.A. (2010). Dopamine and adaptive memory. Trends Cogn Sci *14*, 464-472.
28. McGaugh, J.L., and Roozendaal, B. (2002). Role of adrenal stress hormones in forming lasting memories in the brain. Curr Opin Neurobiol *12*, 205-210.
29. Roig, M., Thomas, R., Mang, C.S., Snow, N.J., Ostadan, F., Boyd, L.A., and Lundbye-Jensen, J. (2016). Time-Dependent Effects of Cardiovascular Exercise on Memory. Exerc Sport Sci Rev *44*, 81-88.
30. Statton, M.A., Encarnacion, M., Celnik, P., and Bastian, A.J. (2015). A Single Bout of Moderate Aerobic Exercise Improves Motor Skill Acquisition. PLoS One *10*, e0141393.
31. Roig, M., Skriver, K., Lundbye-Jensen, J., Kiens, B., and Nielsen, J.B. (2012). A single bout of exercise improves motor memory. PLoS One *7*, e44594.
32. Rhee, J., Chen, J., Riechman, S.M., Handa, A., Bhatia, S., and Wright, D.L. (2015). An acute bout of aerobic exercise can protect immediate offline motor sequence gains. Psychol Res.
33. Nanda, B., Balde, J., and Manjunatha, S. (2013). The Acute Effects of a Single Bout of Moderate-intensity Aerobic Exercise on Cognitive Functions in Healthy Adult Males. J Clin Diagn Res *7*, 1883-1885.
34. Basso, J.C., Shang, A., Elman, M., Karmouta, R., and Suzuki, W.A. (2015). Acute Exercise Improves Prefrontal Cortex but not Hippocampal Function in Healthy Adults. J Int Neuropsychol Soc *21*, 791-801.

Figure Legends

Figure 1. Experimental design. IE=Immediate Exercise group. DE=Delayed Exercise group. NE=No Exercise group. Participants first learned 90 picture-location associations and completed a baseline memory test (Test 1) before undergoing two control sessions (NE), or a control and a exercise session (IE and DE) separated by a 3hr delay period. 48hrs about initial learning, a delayed memory test (Test 2) was administered in the magnetic resonance scanner. See also **Figure S1** for a visualization of the experimental trials and **S2** for the exercise protocol and measures of heart rate, exercise load and ratings of exercise intensity.

Figure 2. Memory retention (Test 2 / Test 1 performance). Error bars denote standard errors of the mean. Asterisks denote significant differences at $p_{TukeyHSD}<0.05$. See also **Figure S3** for raw performance scores.

Figure 3. BOLD activation during Test 2. A. BOLD Contrast between correct trials and null events.
B. BOLD Contrast between correct trials and incorrect trials. Results from these group analyses include the full sample of participants. Activation shown on the SPM8 MNI template at $p_{voxel}<0.001$ and $p_{cluster}<0.05$(Family-Wise Error corrected). MNI coordinates are given for the relevant axis. See also **Table S3** for statistics.

Figure 4. Hippocampal pattern similarity during Test 2. Thresholds are based on the 95th percentile of the computed distribution of random permutations; similarity above this threshold is higher than expected by chance (given a p_{chance} of 0.05). Error bars denote standard errors of the mean. L= Left, R= Right. See also **Figure S4** for pattern similarity in other ROIs.

Supplemental Experimental Procedures

Participants

72 healthy participants (mean age ± standard deviation: 21.93±2.5 years, range 18-28; 48 females) were included in this study and performed all procedures and tests. Participants were recruited through a local university database; all participants were students from tertiary educational institutes. To be considered for participation, participants had to be healthy, between 18-28 years old, perform physical exercise regularly (1-5 times a week) and have a Body Mass Index (BMI) between 18.5 and 25 kg/m2 (mean BMI ± standard deviation: 21.7±1.7 kg/m2, range 18.6-25.0; no significant BMI differences between groups; see **Table S1**). For this reason, participants were medically screened before inclusion. This screening included blood pressure, heart rate and BMI measures as well as the recording of a resting electrocardiogram (ECG) for participants in both exercise groups. Additionally, a sports-medical interview was conducted detailing each participant's medical history and current physical activity. Exercise habits (weekly exercise frequency and duration) did not differ between groups (see **Table S1**). This finding would suggest that fitness levels were not significantly different between groups, but we did not explicitly measure these using e.g. a VO_2 max test. Exclusion criteria were usage of medication (except paracetamol and oral contraceptives), inability to perform bicycle exercise, recent illness (in the two weeks before the experiment), hypertension (systolic blood pressure >140 mmHg or diastolic blood pressure >90 mmHg), any current or past cardiovascular abnormality or any abnormality found on the ECG during screening, any current or past neurological disorder, diagnosed diabetes mellitus, diagnosed hypercholesterolemia, smoking or having quit smoking less than 2 years ago, and any family history of death before the age of 50 due to a cardiovascular disorder. Participants that were magnetic resonance imaging (MRI) incompatible were excluded as well. In total, 95 participants were initially considered for participation. Of those, 10 were excluded during the screening process for not meeting the health or sport criteria listed in the methods section. After passing the screening, 7 participants could not be scheduled for various reasons (recent or current illness, limited availability, etc.), leaving a total of 78 participants that were admitted to the testing phase of the experiment. For 4 participants, MRI data could not be acquired due to technical problems with the MR scanner. 1 participant withdrew from the study at the onset of scanning due to claustrophobia, and 1 participant withdrew from the study due to illness (unrelated to performing exercise). In the end, 72 participants successfully completed all tests and procedures. No participants were excluded due to head motion. The study was approved by the local ethics committee (CMO Region Arnhem-Nijmegen, the Netherlands). All study procedures were in accordance with national legislation for the protection of human volunteers in non-clinical research settings and the Helsinki Declaration [S1]. Written informed consent was obtained before participation. Participants received 76 Euros reimbursement after full participation.

Procedures

After screening and inclusion, participants were randomly assigned to one of three experimental groups: the No Exercise group (NE), the Immediate Exercise group (IE), or the Delayed Exercise group (DE). The obvious nature of the exercise intervention made blinding impossible (for both experimenter and participant). Since the analyses were conducted by the same individuals who acquired the data, this also meant that analysts were not blind to group membership. The proportion of males and females was not significantly different between experimental groups (NE: 17/24 female, IE: 15/24 female, DE: 16/24 female; Pearson Chi-Square Test: χ^2=0.375, p=0.829). Age also did not differ between groups (One-way ANOVA, F_{Age}=1.415, p=0.25). Because circadian influences on memory have been observed previously [S2, S3], the three experimental groups were further subdivided into a morning (AM, starting at 9AM) and an afternoon (PM, starting at 12PM) group to check for time of day effects. We did not find any evidence that time-of-day had direct effects on memory performance or interacted with the observed effects of exercise, as mentioned in the Results section. The experiment took place on two separate days. On the first day of the experiment, participants performed an associative memory task (an extensive description of the task procedures can be found below). Participants' memory performance was tested immediately after learning, providing a baseline performance score (*Test 1*). Participants subsequently underwent the first of two experimental sessions, which was an exercise session for the IE group and a control session for the DE and NE groups (details on both experimental sessions can be found below). After this session, participants were moved to a relaxing living room-like environment to watch nature documentaries (Life, BBC, 2005; Planet Earth, BBC, 2006; Earthflight, John Downer Productions Ltd, 2011; Human Planet, BBC, 2011) for three hours (the *delay period*). The second experimental session was performed approximately 4 hours after learning, and consisted of an exercise session for the DE groups and a control session for the IE and NE groups. The memory task (both learning as well as Test 1) and the two

experimental sessions were always performed in the same room. On the second experimental day, 48±1 hours after learning, participants performed the second memory test (*Test 2*). Test 2 was performed in a magnetic resonance scanner while functional magnetic resonance imaging (fMRI) data was recorded. A structural scan was obtained afterwards and was used during fMRI preprocessing for normalization purposes. Participants were instructed to refrain from doing sports between leaving the lab on Day 1 until their return to the lab on Day 2.

Memory task

Participants intentionally studied 90 picture-location associations in two rounds of passive memory encoding separated by one round of active memory retrieval behind a PC. Picture stimuli consisted of 90 color pictures of round real-life objects (balls, plates, bowls, etc.). All pictures were identical in size. Each of the 90 pictures was associated with one of six different locations. Locations were arranged in a circle on a black background (see **Figure 1**) and were equidistant from the centre of the computer screen. During memory encoding, participants observed passively as pictures were shown one by one and moved to their correct locations on the computer screen. Each picture was first shown in the centre of the screen for 1000 ms. Subsequently, its correct location was highlighted for 1000 ms before the picture was moved to this location in 1000 ms. The picture would then be displayed at its correct location for 2000 ms. The trial concluded with a 1000 ms inter-trial interval (ITI), during which only a white fixation cross was shown. During memory retrieval, participants practiced recalling the picture-location associations. Participants were cued by the presentation of the picture in the centre of the screen and subsequently used a joystick (Logitech Attack 2, Logitech, Lausanne, Switzerland) to move to and select the location they believed to be correct. Selection was confirmed with a pull on the joystick trigger. There was no time limit during retrieval; trials terminated when the participant had selected one of the locations. Trials were separated by an ITI of 3000 ms. No immediate feedback was given; however, participants were shown the total number of objects placed correctly at the end of the retrieval phase. Trial presentation order was randomized for all task phases and for each participant separately.

Directly after study, the first cued-recall memory test (*Test 1*) was conducted, which served as a baseline test for memory performance. During Test 1, participants were cued with the picture and subsequently had to place it on its correct location using the joystick. Participants had 5000 ms per trial to make their response. If no response was given by the participant within 5000 ms, the trial was scored as a no response trial. After choosing a location, and a 1000 ms interstimulus interval (ISI), participants indicated their confidence in their choice on a color scale (1-5, with 1=unsure and 5=very sure), for which there was no time limit. Trials were separated by an ITI of 2000 ms. A second test (*Test 2*) was administered 48 hours after study and was performed inside the MR scanner. Test 2 was mostly identical to Test 1, but used a variable ITI (jittered between 3000-7000 ms on each trial) and also included 15 null events. Null events were trials in which the recollection procedure (cued recall plus confidence rating) was replaced with a white fixation cross for 10000 ms. These events were randomly introduced in the test and served as an explicit baseline for recall-related events (i.e. they allowed for a BOLD fMRI contrast between task and no task conditions).

Experimental sessions

Participants underwent two experimental sessions. IE and DE participants experienced both an exercise session and a control session; NE participants took part in two control sessions. Both sessions took place in the same room and lasted 35 minutes. The exercise session consisted of a single, sub-maximal interval training on a stationary bicycle ergometer (Optibike 50 med, ergoline GmbH, Bitz, Germany). The characteristics of our exercise session were based on findings from [S4, S5] and advice given by colleagues from the Radboudumc Physiology Department (Maria Hopman, personal communication). Load was continuously adjusted to the participants' heart rate. During the training, exercise intensity went up to 80% of the participants' maximum heart rate; the intensity graph of the full training protocol can be seen in **Figure S2**. Briefly, the training started with a 5 minute warm-up period during which the exercise load was gradually increased until the participants reached 80% of their maximum heart rate. This heart rate was maintained during the subsequent high intensity interval for 4 minutes (through adjustment of the exercise load if necessary). A 3 minute low intensity interval followed during which load was reduced and participants' heart rate slowly decreased naturally. This high-low intensity interval sequence was performed four times in total. Each high intensity interval after the first started with a ramp-up period during which the load was substantially increased to reach 80% of the participants' maximum heart rate within 1 minute. The training concluded with a 5 minute cooldown period during which the exercise load was gradually reduced. The age-adjusted

maximum heart rate was calculated for each participant using Tanaka's formula [S6]. While exercising, participants were shown a nature documentary (Planet Earth, BBC, 2006); two particular documentaries from this series were chosen and displayed in a counterbalanced manner between participants for control and exercise sessions. During the control session, participants did not perform exercise but instead relaxed and watched the nature documentary from a desk chair. During both exercise and control sessions, heart rate was continuously recorded with a wireless chest band (ergoline GmbH, Bitz, Germany) using a recording frequency of 0.08 Hz. During the exercise session, load was continuously recorded at 0.08 Hz by the ergometer. During both sessions, subjective intensity scores were repeatedly probed (indicated by "B" in **Figure S1**) with the Borg's Rating of Perceived Exertion (Borg-RPE) scale [S7]. These subjective ratings of exercise intensity used a scale of 0 to 10, with 10 being the highest level of intensity ever experienced or imaginable. During both exercise and control sessions, participants were instructed to watch the movies and report their subjective intensity ratings when prompted by the experimenter.

Behavioral analyses

Behavioral analyses were performed using SPSS 21 (IBM Corporation, Armonk, USA). Comparisons between the experimental groups regarding memory performance, reaction times and confidence ratings were made using univariate linear models that include Group (IE/DE/NE), Gender (male/female) and Starting time (AM/PM) as fixed factors. Post-hoc pair-wise analyses of significant effects were conducted using Tukey Honest Significant Difference tests to control for multiple comparisons.

Comparisons between subjective intensity ratings for the exercise and control sessions were made using Repeated Measures general linear models that included a factor Group (IE/DE/NE) and a within-subjects factor Time (timepoint B1-B4). For the control session, post-hoc pair-wise tests were run using Sidak corrections to control for multiple comparisons. Heart rate and load during the exercise session were compared between IE and DE at each sampling point (n=175) using Independent Samples T-tests at a threshold of p=0.001 (uncorrected for multiple comparisons).

MR Data acquisition

MRI data were acquired using a 3.0 T Siemens Skyra (Siemens Medical, Erlangen, Germany) with a multi-echo sequence and a 32 channel head coil system at the Donders Institute, Centre for Cognitive Neuroimaging in Nijmegen, the Netherlands. A 2D Echo Planar Imaging (EPI) multi-echo sequence was used with the following imaging parameters: 31 ascending slices, voxel size = 3.5 x 3.5 x 3 mm, repetition time (TR) = 2390 ms, flip angle α =90°, echo time (TE) = 9.4/20.6/32/43/54 ms, number of echoes = 5. A 3D magnetization-prepared rapid gradient echo (MPRAGE) anatomical T1-weighted image was also acquired with the following parameters: 192 slices, 1.0 mm isotropic voxel size, TR = 2300 ms, TE = 3.0 ms, flip angle α = 8°. Total scanning time was approximately 45 minutes.

fMRI data preprocessing

Preprocessing of fMRI images was done using SPM8 (fil.ion.ucl.ac.uk/spm/software/spm8) and included multi-echo combination, slice-timing correction, co-registration, segmentation, normalisation and smoothing using default SPM parameters. Multi-echo combination was done using in-house developed software and included realignment with SPM. Slice-timing correction was performed using the middle slice as reference. Realigned, slice-time corrected EPI images were coregistered to participant anatomical T1 images. Normalisation was performed by first normalizing individual T1 structural images to the SPM8 T1 template and subsequently applying the resulting transformations to the realigned, coregistered EPI images. Smoothing was then done using an isotropic Gaussian kernel with a full-width at half-maximum (FWHM) of 8 mm [S8, S9]. Segmentation of participant T1 anatomical images into grey matter, white matter (WM) and cerebrospinal fluid (CSF) segments was used as a basis for nuisance regression; mean signal intensity from WM and CSF compartments was calculated from each functional image using in-house procedures and subsequently used to create two nuisance regressors that were included at the first level analysis for every participant [S10].

First and second level analysis

After preprocessing, a contrast image reflecting task-related brain activity (i.e. Blood Oxygenation Level Dependent (BOLD) signal differences between correct trials and null events) was estimated for each participant with the general linear model (GLM). In addition, a contrast image reflecting correct memory retrieval (i.e. BOLD signal differences between correct and incorrect trials) was estimated. The GLM used an event-related design matrix in which types of events were estimated in individual regressors. Separate regressors modeled the onsets of correct responses, incorrect responses, missed responses (when no response was given within 5 seconds), confidence ratings, ITIs and null events during the task. Of these events, confidence ratings and ITIs were modeled as stick-functions with a duration of 0. Missed responses were assigned a duration of 5 seconds, and null responses a duration of 10 seconds. Correct and incorrect responses were modeled with a duration as long as their corresponding reaction times, i.e. the time between the stimulus onset and the participant's response. The BOLD impulse response was modeled using the canonical hemodynamic response (HRF) in SPM and also included time derivatives for each regressor. Additionally, the realignment parameters derived from the realignment preprocessing step and their first derivatives were included to partially correct for the participant's movement during scanning. Finally, the two compartment signals were added as additional nuisance regressors to correct for residual movement and artifacts. The first-level contrast images were subsequently entered into second level group analyses using full-factorial models that included the factors Group (IE, DE and NE) and Condition (Correct vs Incorrect and Correct vs Null, respectively). Differences between experimental groups were calculated using F-contrasts and family wise error (FWE) corrected at the voxel level at $p<0.05$. In the absence of differences between experimental groups, effects of correct recall (Correct>Incorrect) and task-related activity (Correct>Baseline) were then estimated across the full sample of participants using T-contrasts and corrected for multiple comparisons at the cluster level at FWE $p<0.05$ after initial correction at the voxel level at $p<0.001$ (uncorrected).

Representational Similarity Analysis

Representational similarity analysis (RSA) [S11] was used to determine neural pattern similarity (PS) within types of trials (i.e. conditions, in this case correct and incorrect responses), and was performed on unsmoothed data within the native-space of each participant. Single-trial estimates of the BOLD response were obtained by modeling each recall trial as a separate regressor in a single design matrix that included also the realignment parameters and their first derivatives. From this design matrix, 90 t-maps were calculated per participant (individually divided between a number of incorrect and correct t-maps based on Test 2 performance). RSA was initially performed on regions-of-interest (ROIs) in the hippocampus, comprising of two ROIs centered on the local maximum of activation in the left and right hippocampus in the 2nd level contrast between correct > incorrect responses (obtained at the $p_{voxel}=0.001$, $p_{clusterFWE}=0.05$ threshold). A 10 mm sphere was centered on these peak voxels. To constrain the ROIs to the anatomical hippocampus, the left and right hippocampal segments from the Automated Anatomical Labeling (AAL) atlas were back-transformed into each participant's native space and the overlap between the resulting mask and the 10mm spheres was used to delimit the two hippocampal ROIs. To compare the hippocampal results with results from other ROIs, we afterwards also created control ROIs from local activation maxima in the bilateral striatum, the bilateral inferior parietal cortex and the medial prefrontal cortex. These local maxima were also obtained from the correct > incorrect contrast, at the same threshold, and the extent of the resulting ROIs was again limited by 10mm spheres. The signal from voxels within each ROI was extracted and used to create a pattern vector. This extraction provided a trial × voxel matrix, in which the individual trials were sorted according to correctness (correct versus incorrect responses). Values were converted into *z*-scores to correct for differences in mean activation between trials and single-trial PS values were then calculated by correlating each trial with all other trials (Pearson's *r*). Single-trial PS values were subsequently transformed into Fisher's *z* values, and mean PS scores for each condition (correct x correct, incorrect x incorrect) were calculated by collapsing across the respective segments of the data matrix. Differences between mean PS scores were tested using non-parametric permutation tests. Specifically, trial labels were shuffled randomly across 10000 permutations. Then, mean PS scores for each randomly permuted condition were calculated as described previously. Participant-specific permuted PS scores were consequently averaged across the entire participant sample. The resulting null-distribution was thresholded at the 95% percentile (thus accepting a type I error of 5%) to determine the cutoff for 'significant' similarity above chance-level for each ROI separately. A single outlier was detected; one participant in the IEG displayed PS values greater than the mean IEG value plus three times the standard deviation for striatal and hippocampal ROIs during correct responses. This person was excluded from the pattern analyses; however, including this person does not significantly change the results. Pattern

similarity was compared between groups using Repeated Measure analyses that included the factors Group (IE/DE/NE), Correctness (Correct versus Incorrect responses) and Hemisphere (Left versus Right, included for all ROIs but the uniquely medial prefrontal cortex ROI). Significant effects and interactions were further explored using pairwise post-hoc comparisons that controlled for multiple comparisons using Sidak corrections.

Actigraphy

Between experimental day 1 and experimental day 2, participants' physical activity was recorded continuously using actigraphy (ActiGraph GT3X, ActiGraph, Pensacola, USA). Total sleeping time (TST) could consequently be estimated for each participant from this data using proprietary Actigraph analysis software (ActiLife, ActiGraph, Pensacola, USA), and was calculated using the Cole-Kripke algorithm [S12]. We found no significant differences between experimental groups for the TST during the first, second or combined nights, and none of the TST measures correlated with memory retention, either across the sample or within the IE, DE or NE separately (all $p>0.05$). It therefore appears that the effects reported in this article are not related to the TST of our participants during the experiment. Average activity, defined as the mean activity count per 5 minute epoch in the period between experimental day 1 and experimental day 2 did not differ between experimental groups ($\mu_{IE} \pm SD = 635.1 \pm 202.7$; $\mu_{DE} = 692.1 \pm 252.7$; $\mu_{NE} = 760.3 \pm 200.3$; One Way ANOVA: $F = 1.880$, $p=0.161$).

Supplemental References

S1. World Medical, A. (2013). World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. JAMA *310*, 2191-2194.

S2. Schmidt, C., Collette, F., Cajochen, C., and Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. Cogn Neuropsychol *24*, 755-789.

S3. Gerstner, J.R., and Yin, J.C. (2010). Circadian rhythms and memory formation. Nat Rev Neurosci *11*, 577-588.

S4. Winter, B., Breitenstein, C., Mooren, F.C., Voelker, K., Fobker, M., Lechtermann, A., Krueger, K., Fromme, A., Korsukewitz, C., Floel, A., et al. (2007). High impact running improves learning. Neurobiol Learn Mem *87*, 597-609.

S5. Roig, M., Nordbrandt, S., Geertsen, S.S., and Nielsen, J.B. (2013). The effects of cardiovascular exercise on human memory: a review with meta-analysis. Neurosci Biobehav Rev *37*, 1645-1666.

S6. Tanaka, H., Monahan, K.D., and Seals, D.R. (2001). Age-predicted maximal heart rate revisited. J Am Coll Cardiol *37*, 153-156.

S7. Borg, G.A. (1974). Perceived exertion. Exerc Sport Sci Rev *2*, 131-153.

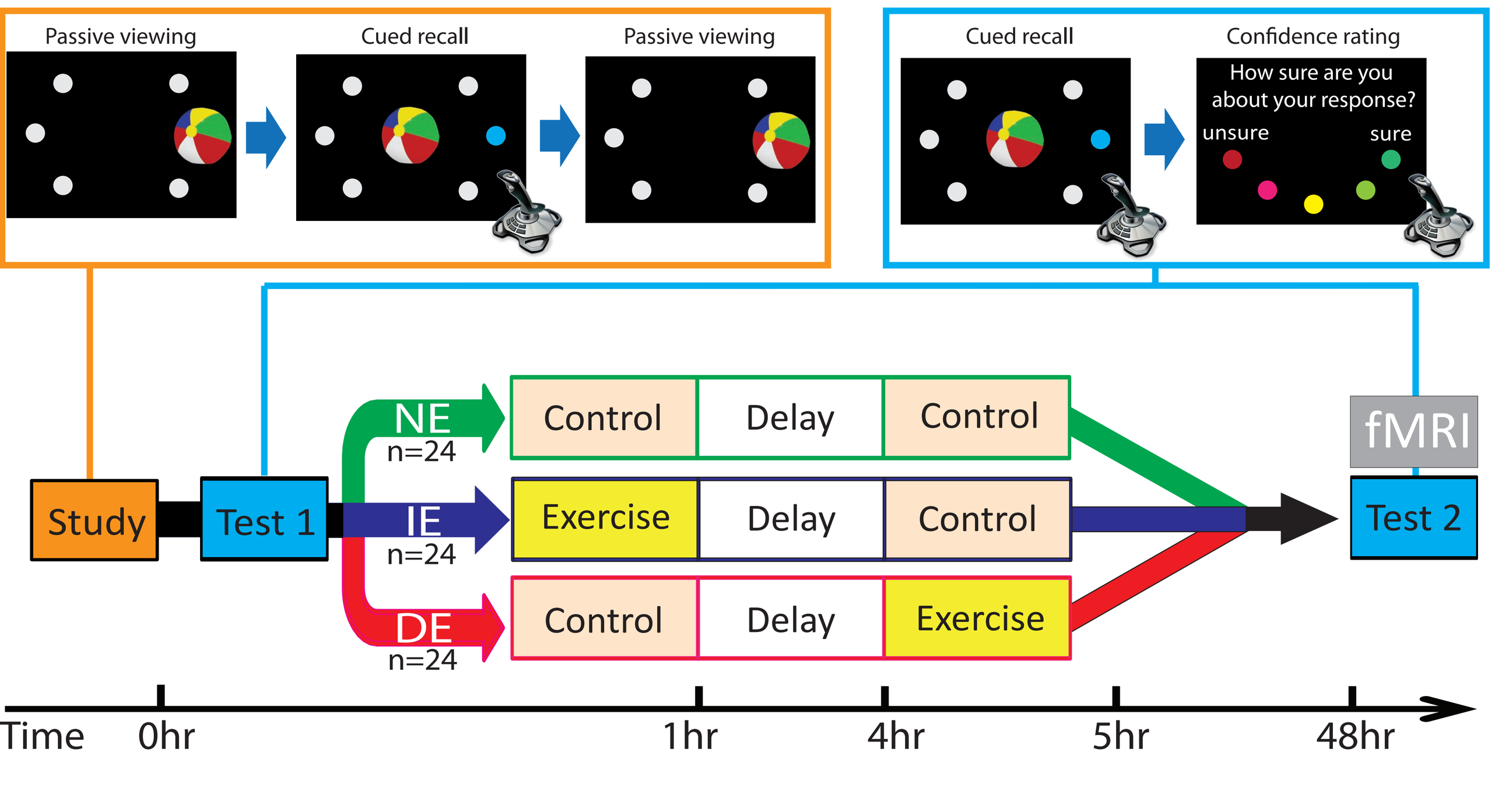
S8. Mikl, M., Marecek, R., Hlustik, P., Pavlicova, M., Drastich, A., Chlebus, P., Brazdil, M., and Krupa, P. (2008). Effects of spatial smoothing on fMRI group inferences. Magn Reson Imaging *26*, 490-503.

S9. Worsley, K.J., and Friston, K.J. (1995). Analysis of fMRI time-series revisited--again. Neuroimage *2*, 173-181.

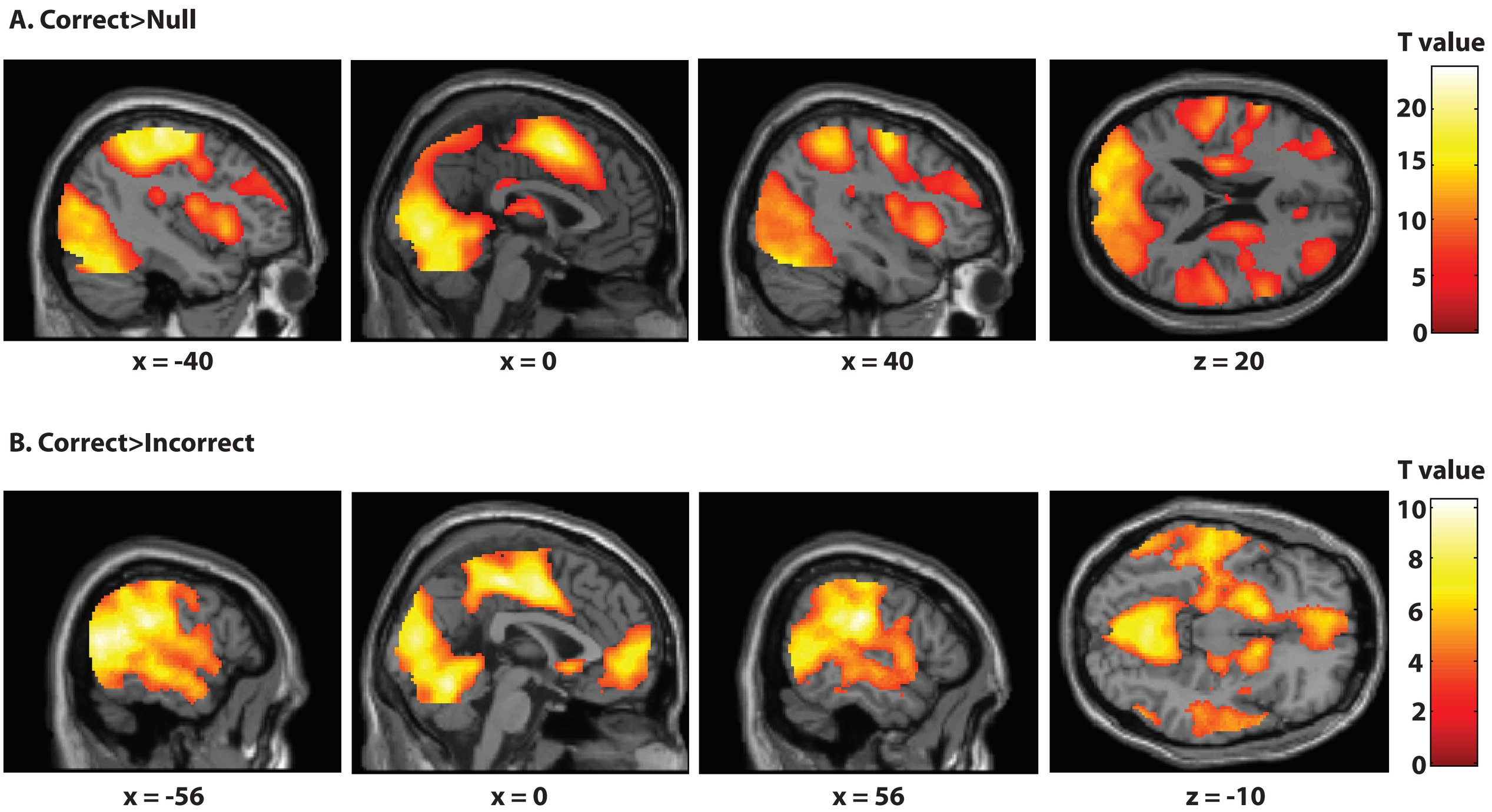
S10. Verhagen, L.G., M.J.; Dijkerman, H.C; Toni, I., (2006). Studying visually guided reach to grasp movements in an MR-environment. Neuroimage *31*, S45.

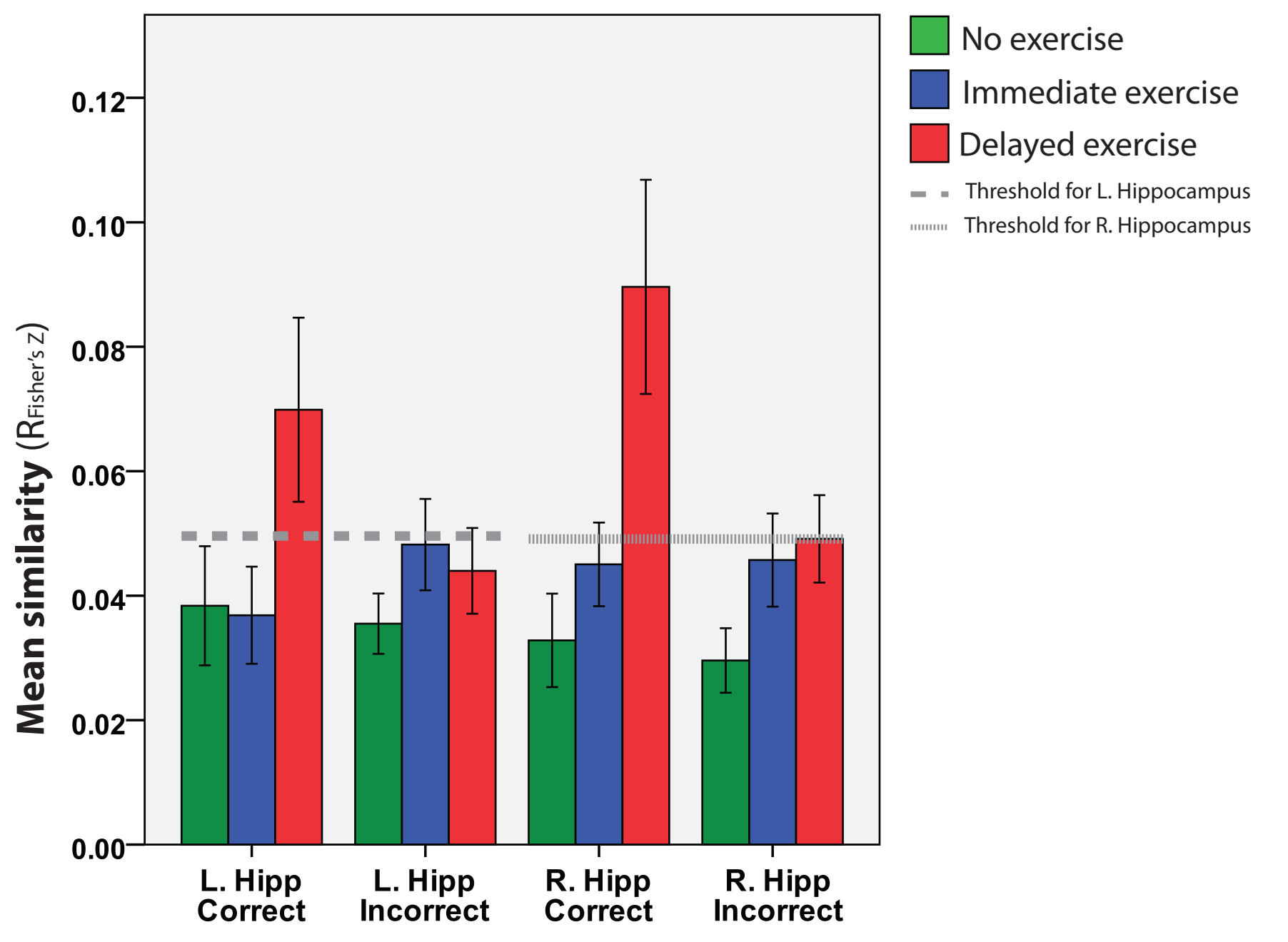
S11. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci *2*, 4.

S12. Cole, R.J., Kripke, D.F., Gruen, W., Mullaney, D.J., and Gillin, J.C. (1992). Automatic sleep/wake identification from wrist activity. Sleep *15*, 461-469.

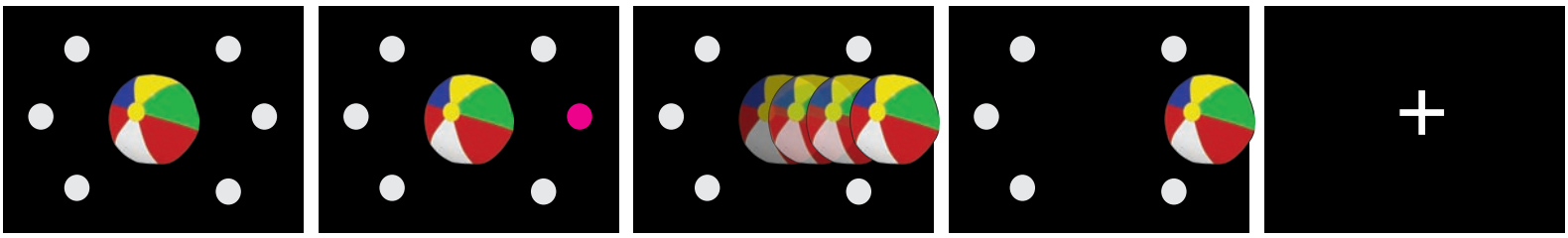

Passive viewing
Cued recall
Passive viewing
Cued recall
Confidence rating
How sure are you about your response?
unsure
sure
Study
Test 1
NE
n=24
IE
n=24
DE
n=24
Control
Delay
Control
Exercise
Delay
Control
Control
Delay
Exercise
fMRI
Test 2
Time
0hr
1hr
4hr
5hr
48hr

No exercise
Immediate exercise
Delayed exercise
*
*
1.00
0.90
0.80
0.70
0.60
0.50
Retention (Test 2/Test 1 performance)


A. Correct>Null
T value
20
15
10
5
0
x = -40
x = 0
x = 40
z = 20
B. Correct>Incorrect
T value
10
8
6
4
2
0
x = -56
x = 0
x = 56
z = -10

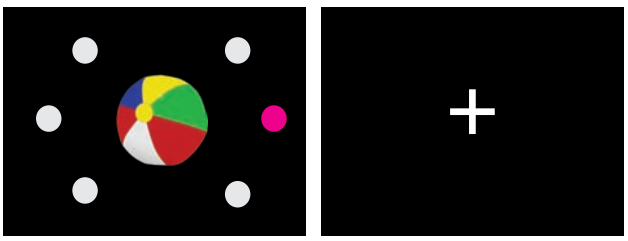

No exercise
Immediate exercise
Delayed exercise
Threshold for L. Hippocampus
Threshold for R. Hippocampus
Mean similarity (RFisher's z)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
L. Hipp Correct
L. Hipp Incorrect
R. Hipp Correct
R. Hipp Incorrect

Passive viewing



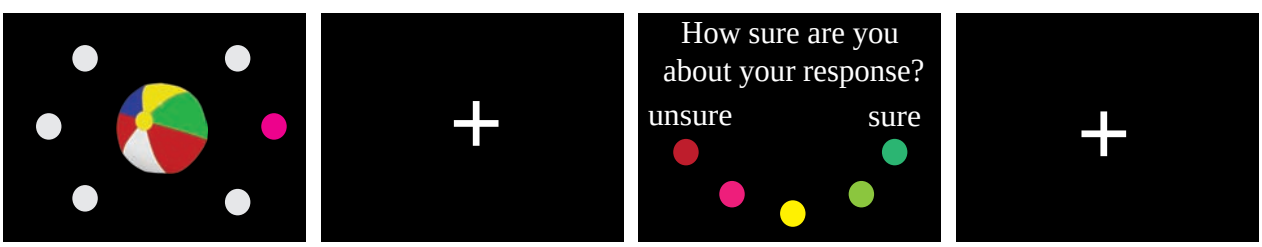
1000ms 1000ms 1000ms 2000ms 1000ms

Cued recall (Study)



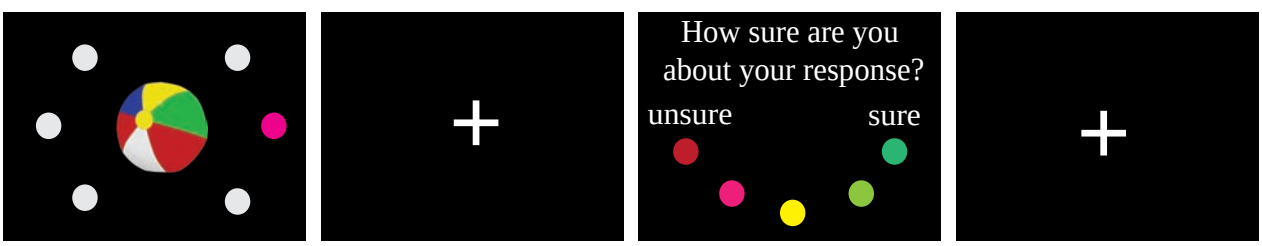
until response 3000ms

Cued recall (Test 1)



max. 5000ms or until response 1000ms until response 2000ms

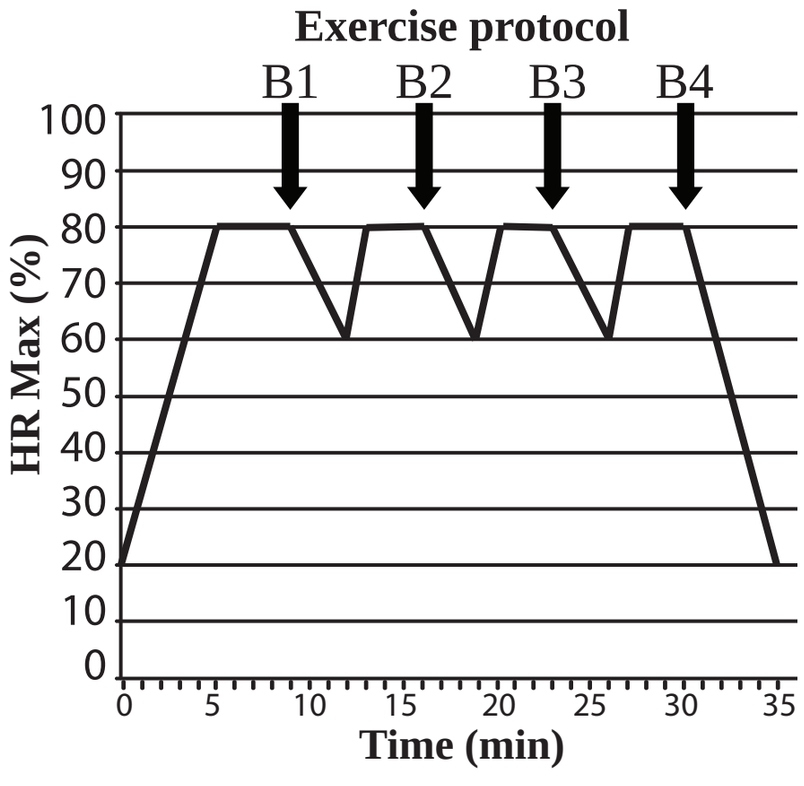
Cued recall (Test 2)

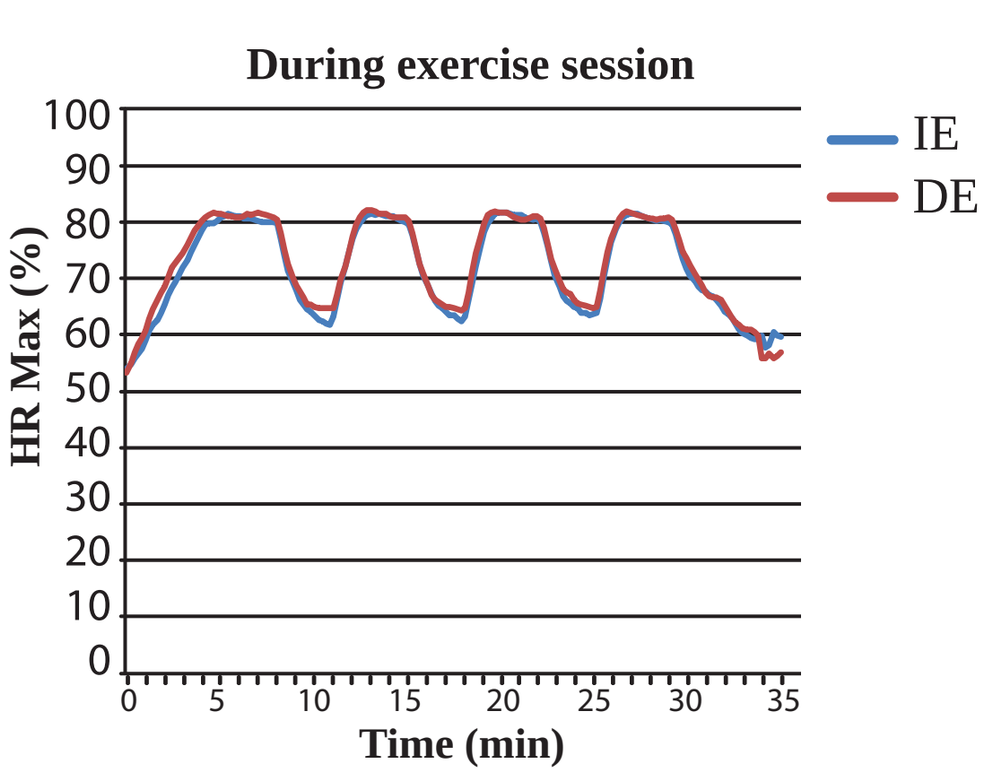


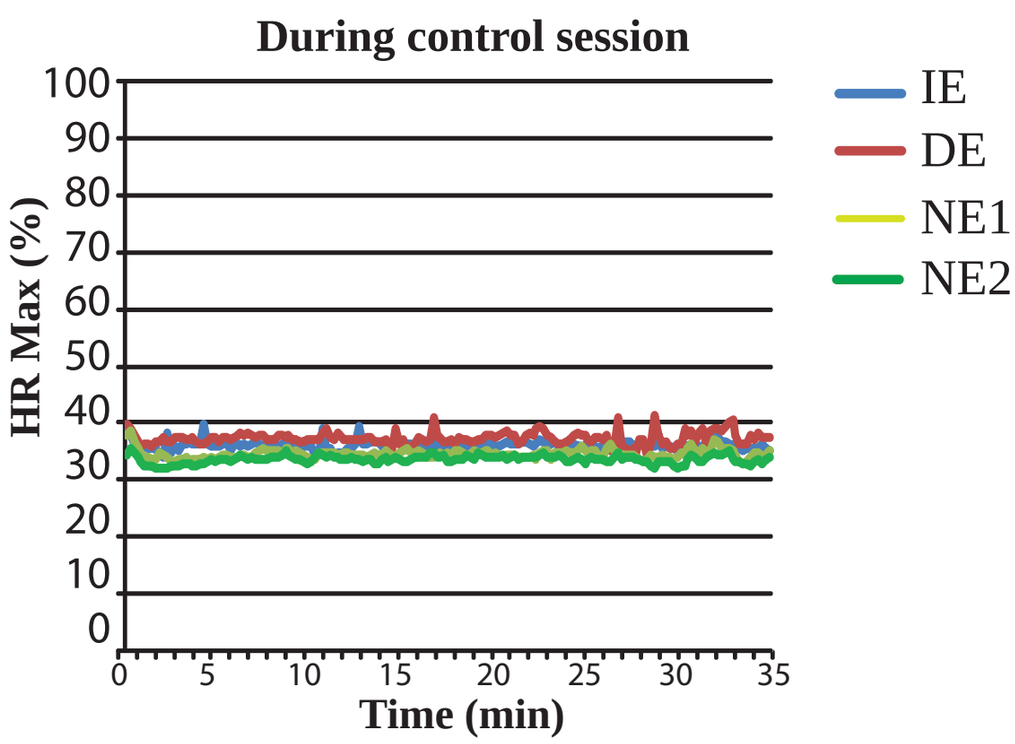
max. 5000ms or until response 1000ms until response 3000-7000ms

Figure S1. Example trials for the different phases of the picture-location associative memory task. Related to Figure 1.

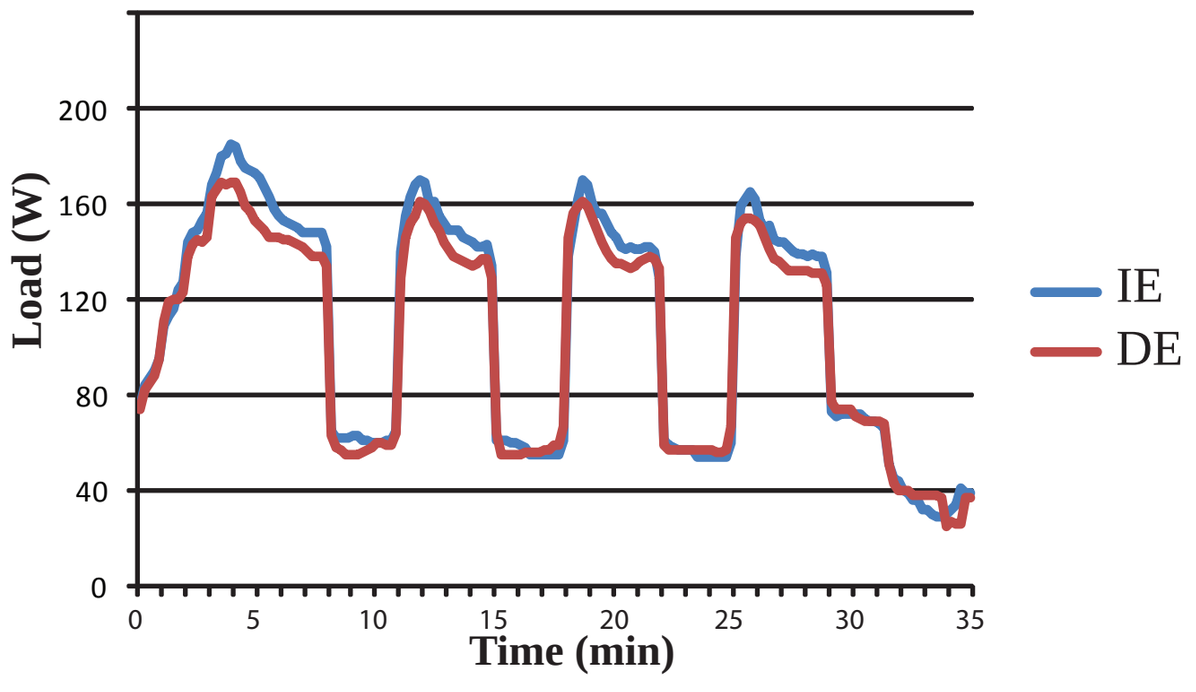
A. Heart rate

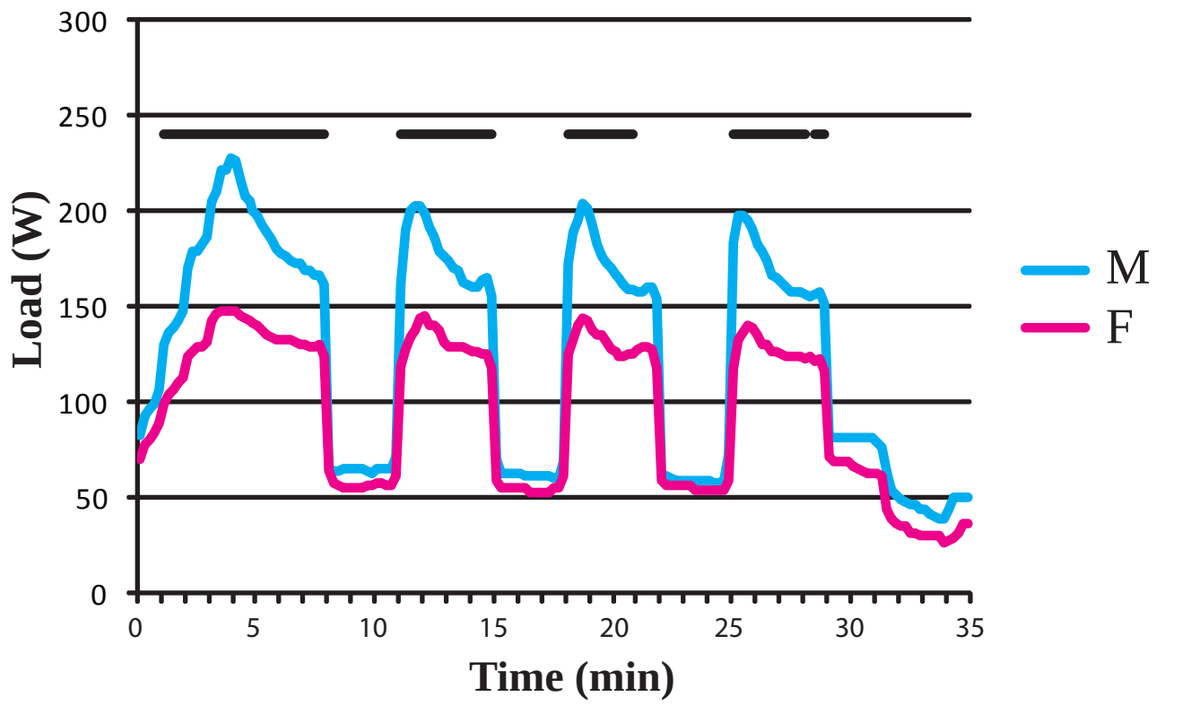






B. Load during exercise session





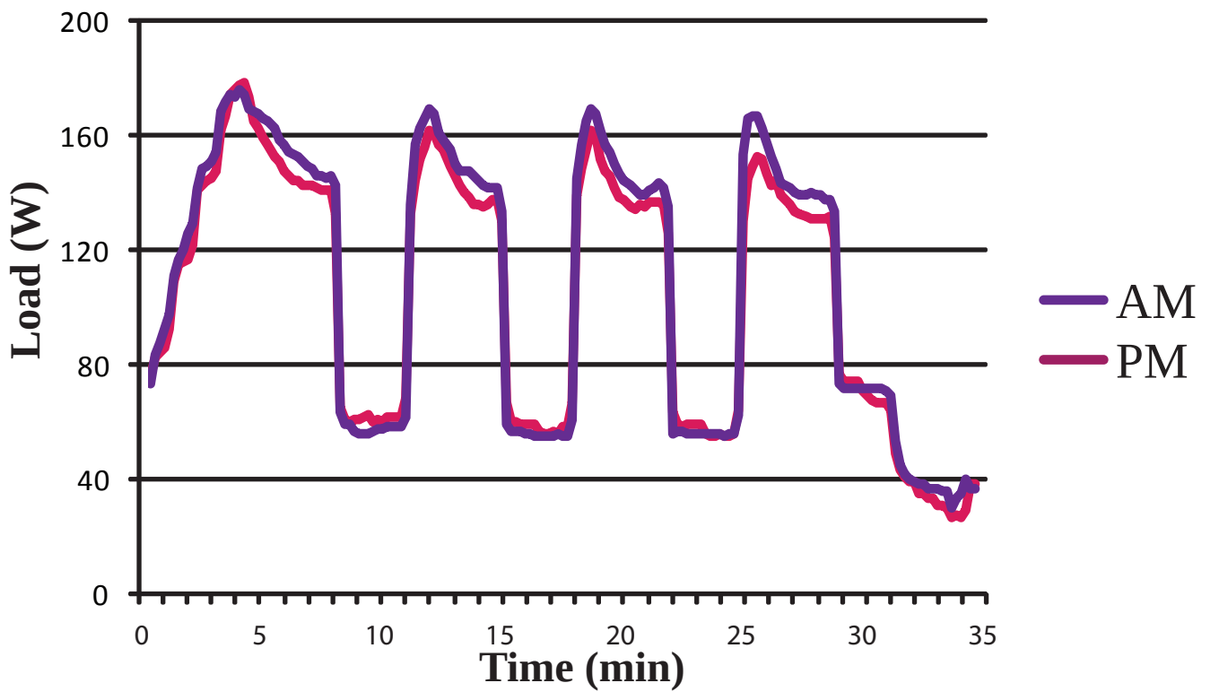


Figure S2. Exercise protocol, heart rate and exercise load
Related to Figure 1.
A. Mean heart rate. Graphs show target heart rate according to protocol, actual heart rate during exercise (for IE/DE) and actual heart rate during control sessions (for all groups).
B. Exercise load during exercise session. M=male, F=female. AM=morning group, PM=afternoon group.
Black bars denote significant differences between groups (at $p<0.001$, uncorrected for multiple comparisons).

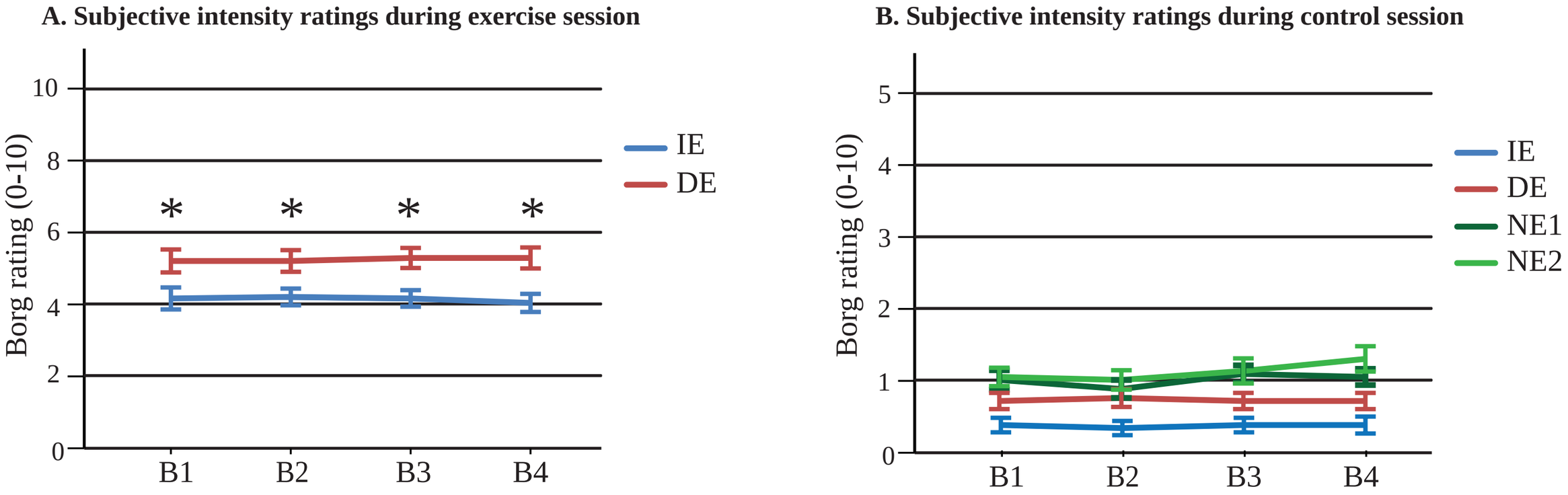


Figure S3. Subjective intensity ratings
Related to Figure 1.

A. Subjective intensity ratings during the exercise session. A Repeated Measures General Linear Model including the factors "Group" and "Time" (i.e., measurement B1-B4) indicated that the IE group rated the exercise intensity significantly lower than the DE group throughout the exercise session (F_{Group}=9.56, p_{Group}=0.003; F_{Time}=0.148, p_{Time}=0.931; $F_{GroupxTime}$=0.268, $p_{GroupxTime}$=0.848). However, absolute measures of intensity (heart rate, exercise load) were not significantly different between groups nor were subjective intensity ratings correlated with memory retention (all $p>0.05$; Pearson correlations were neither significant across the sample, nor when investigating the three experimental groups separately). One possible reason for the observed difference could be that DE participants experienced a greater change in activity when performing exercise, coming from the relatively passive delay period. Alternatively, DE participants might show a time-on-task effect, having spent already many hours in the lab before exercising, versus the IE group, whose participants exercised relatively early on in the experimental procedures.

B. Subjective intensity ratings during the control session. There was a significant interaction between Group and Time during the control session (F_{Group}=11.80, p_{Group}<0.001; F_{Time}=1.730, p_{Time}=0.166; $F_{GroupxTime}$=2.318, $p_{GroupxTime}$=0.035). Post-hoc pairwise tests using Sidak corrections for multiple comparisons showed that the NE group rated the intensity higher than the IE group at all timepoints (p_{B1}<0.001; p_{B2}<0.001; p_{B3}<0.001; p_{B4}<0.001) but only higher than the DE group at timepoint B3 and B4 (p_{B1}=0.100; p_{B2}=0.505; p_{B3}=0.041; p_{B4}=0.017). These group differences might be due to a bias in the exercise groups: IE and DE participants might have rated the intensity of the control session lower due to their experience or anticipation of a more intense exercise session. NE group participants did not perform any exercise and thus could not show such effects.

NE1/2: control session 1/2 for the no exercise group.
B1-B4 represent the 4 subjective intensity measurements as indicated in Figure S3A; Asterisks denote significant differences.

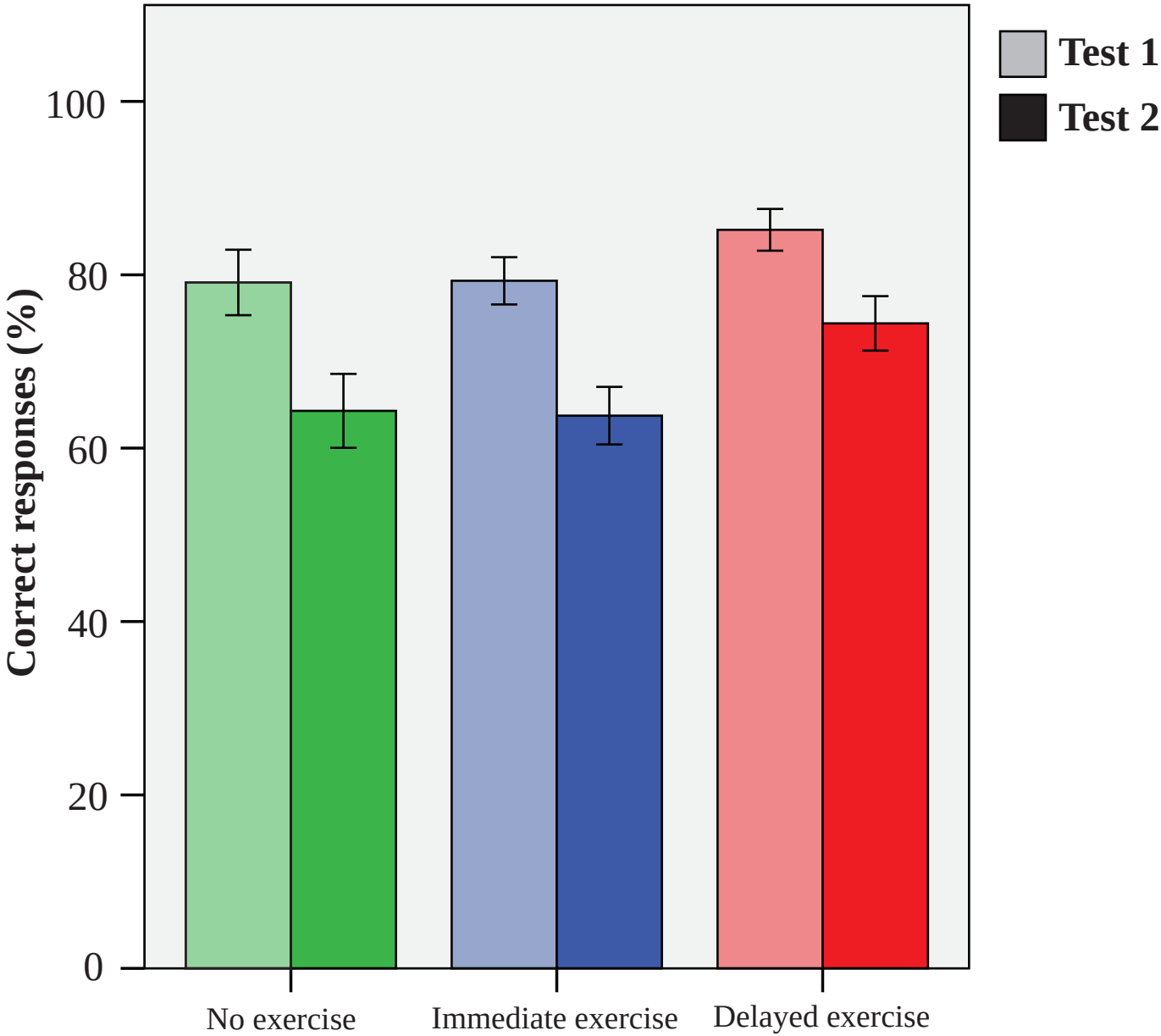


Figure S4. Performance on Test 1 and 2.
Related to Figure 2.
Performance is displayed as the percentage of correctly recalled picture-location associations.
Error bars denote the standard error of the mean.

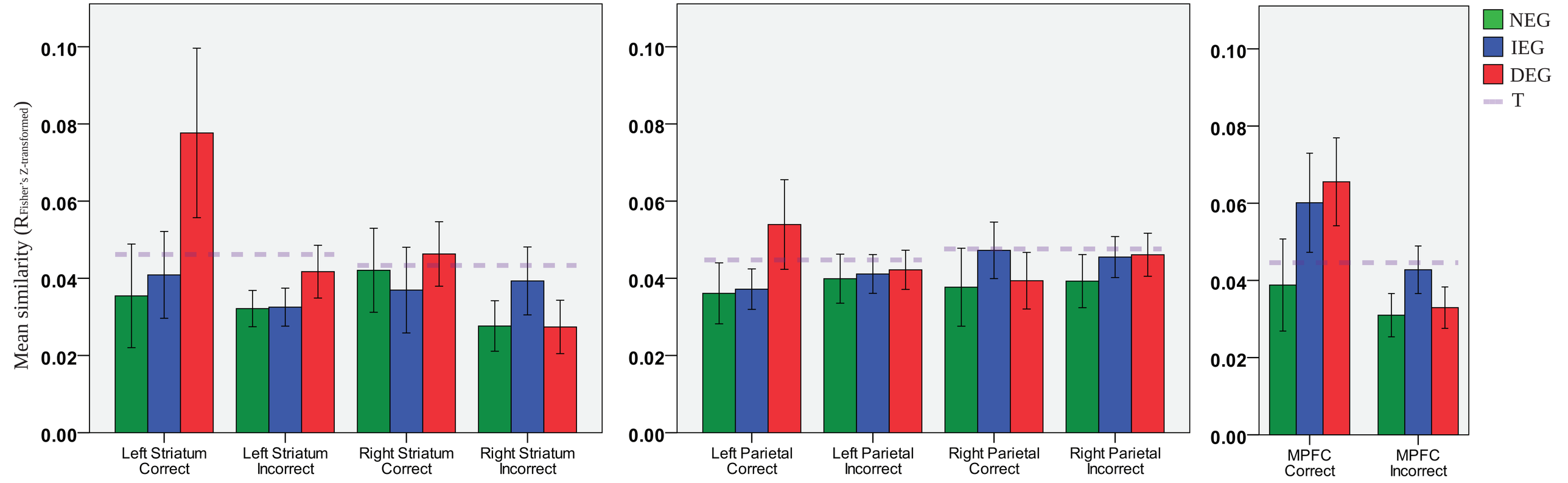


Figure S5. Pattern similarity in other brain regions. Repeated measures ANOVAs with the factors Group, Correctness and Hemisphere (for striatum and parietal ROIs) showed a main effect of Correctness for striatal ($F_{Correct}$=5.74, $p_{Correct}$=0.019) and MPFC ($F_{Correct}$=9.92, $p_{Correct}$=0.002) but not parietal ($F_{Correct}$=0.017, $p_{Correct}$=0.898) ROIs. No significant group or hemispheric differences or any two-way or three-way interactions were observed. T: Threshold based on the 95th percentile of the computed distribution of random permutations; similarity above this threshold is higher than expected by chance (given a p_{chance} of 0.05). Related to Figure 4.