www.sciencemag.org/cgi/content/full/320/5882/1510/DC1

Supporting Online Material for

Transfer of Learning After Updating Training Mediated by the Striatum

Erika Dahlin,* Anna Stigsdotter Neely, Anne Larsson, Lars Bäckman, Lars Nyberg

*To whom correspondence should be addressed. E-mail: erika.dahlin@physiol.umu.se

Published 13 June 2008, *Science* **320**, 1510 (2008)
DOI: 10.1126/science.1155466

This PDF file includes:

Materials and Methods
Fig. S1
Tables S1 to S3
References

Supporting Online Material

MATERIALS AND METHODS

Experiment 1

Subjects

24 younger participants were randomized to training or control groups. The control group made it possible to dissociate test-retest changes from true training-related changes. This is a critical but rarely considered factor in research on the neural correlates of cognitive training. Participants were recruited through e-mail to students at Umeå University, Sweden. All participants who completed the study were paid SEK 1000 (approximately 160 USD). One participant declined to continue participation after the pre-training session and one was identified as an outlier in the analyses of the behavioral data. After exclusion of these individuals, the participants ranged between 20 and 31 years of age (Table S1). The training and control groups did not differ significantly ($p > .05$) in age, years of education, depression (*1*), vocabulary (*2*), mental status (*3*) or mental speed measured by pattern comparisons (*4*) and number copying (*4*). All participants reported being in good physical and mental health, not color blind, and right handed.

Procedure

The training program was a computerized in-house developed visual basic program and consisted of the letter memory criterion task and five other updating tasks. Participants in the training group undertook 5 weeks of training in groups of four. There were three sessions/week, each lasting 45 minutes. All participants were MRI scanned immediately pre- and post this 5-week period. The control group did not receive any training or specific activity during the period between the two MRI sessions.

Training tasks

Letter memory (*5*) constituted the criterion task and was included during both fMRI sessions and in all training sessions. This task involved 10 lists of varied length (between 5-11 letters). Each list consisted of serially presented letters (A-D, 2.0 s per letter) and the task was to recall the four last presented letters as quickly as possible when the presentation list ended. In responding, participants used their right hand: index finger = A; middle finger = B; ring finger = C; pinky finger = D.

The training sessions included five other tasks, all requiring updating. Four of these tasks were similar to the criterion task (letter memory) and involved updating of single items, but to foster generality of updating training they involved different kinds of stimuli (i.e. numbers, letters, colors, and spatial locations). Five lists of items were randomly presented and the task was to recall the four last presented items. For the criterion letter task and the other four updating training tasks, number of correctly recalled 4-item sequences served as outcome measure. Across the 5-week training session, the lists varied in length to manipulate level of difficulty and thereby ensure that the training was sufficiently demanding (low level = 4-7 items; medium level = 6-11 items; high level = 5-15 items). Performance was monitored and level of difficulty was adjusted when the participant scored at 80% or higher in the letter memory training task. All subjects reached the most difficult level at the end of training.

The final training task was a keep-track task. This task, too, taxes updating (*5*), but is structurally quite different from the other updating tasks in the battery of training tasks. Thereby, the inclusion of the keep-track task should contribute toward strengthening of a general non-task specific updating ability. In each trial of this task, 15 words from different semantic categories were presented serially in random order (2.0 s per word) and participants were instructed to mentally place the words into categories (animals, articles of clothing, countries, relatives, sports, professions) indicated by boxes at the bottom of the screen. They had to continuously update the content for each category and remember the last presented word in each category at the end of the presentation. Participants responded by typing the last presented word under each category box when the trial ended. During each training session the participants performed three trials with five target categories, and three with three (low level), four (medium level), or five (high level) target categories depending on the level of training.

fMRI tasks

The fMRI scanning sessions involved three different tasks: The letter memory criterion task, the *n*-back transfer task, and the Stroop transfer task. Participants received instructions and practice on the tasks before entering the scanner. During scanning, they viewed stimuli on a screen via a mirror mounted on a head coil. Stimuli were presented in white on a black background using E-prime 1.1, which also recorded behavioral performance.

The letter memory criterion task consisted of 10 lists of serially presented letters (A-D). Each letter was presented for 2.0 s and between each letter a cross was presented for 1.0 s. The lists were of the following lengths: [7, 7, 9, 9, 11, 13, 9, 15, 13, 15]. After each list, subjects responded by using the right hand: index finger = A; middle finger = B; ring finger = C; pinky finger = D. The maximum response time was 6.0 s and between each list a baseline circle was presented for 21.0 s. Number of correctly recalled 4-letter sequences was used as the dependent measure and the maximum score was 10.

The *n*-back transfer task consisted of 27 lists: nine 1-back lists, nine 2-back lists and nine 3-back lists in random order (*6*). Each list consisted of 10 numbers and each number was presented for 1.5 s with a cross presented for 0.5 s between each item. Thus, different stimuli were included in the criterion task (letters) and the n-back transfer task (numbers). A baseline cross was presented for 20.0 s after every 9th list. Subjects indicated whether or not each item in the list matched an item that occurred one, two, or three items back. To respond, participants used their right hand (index finger = "Yes"; middle finger = "No"). The number of correct Yes-responses was used as the dependent measure. As there were four possible correct responses per list, the maximum score across lists was 36.

The Stroop task consisted of 72 items: 24 congruent words, 24 incongruent words and 24 neutral words, presented in random order. Each item was presented for 3.8 s followed by a fixation cross, with the inter-stimulus interval varying between 0.2 s and 5.2 s. The task was to decide which color each word was written in as quickly as possible. Two squares with different colors, one to the left and one to the right side of the screen, were presented and to respond the participants used either their right or left hand to press a button that corresponded to the correct color. The total scanning time was about 70 minutes.

fMRI data acquisition

Images were acquired using a 1.5 T scanner. Functional T2*-weighted images were obtained with a single-shot gradient echo EPI sequence used for blood oxygen level dependent imaging. The sequence had the following parameters: repetition time: 3000 ms (33 slices acquired), echo time: 50 ms, flip angle: 90°, field of view: 22 x 22 cm, 64 x 64 matrix, and 4.4 mm slice thickness. Five dummy scans were performed prior to the image acquisition to eliminate signals arising from progressive saturation. After acquisition, the images were transferred to a PC and converted to Analyze format.

Analyses of cognitive performance

The results from the pre- and post-training fMRI sessions for letter memory, *n*-back, and Stroop (RT) were analyzed using 2 (Group: trained, control) x 2 (Session: pre-training, post-training) repeated measure ANOVAs. To compare the magnitude of gains, the effect size (the mean performance gain in each group divided by the pooled SD at pre-training) was calculated and differences between training and control groups were tested by two-tailed *t*-tests.

Analyses of brain activity

Functional images were analyzed with Statistical Parametric Mapping Software (SPM2) implemented in Matlab 7.1. All images were corrected for slice timing, realigned and unwarped, normalized to standard anatomical space defined by the MNI atlas (SPM2), and smoothed using an isotropic 8.0 mm FWHM Gaussian filter kernel. Statistical analyses were performed on a voxel-by-voxel basis by modeling the active and baseline conditions for the different tasks, convolved with the canonical hemodynamic response function. A fixation baseline was used in the analyses of both letter memory and *n*-back to make the tasks comparable. In the Stroop task, the congruent condition served as baseline for the incongruent condition. For letter memory, only the updating phase and not the final response phase was modeled for each block. Applying the general linear model to the data resulted in least-square estimates for each participant. The individual contrast images were then used for random-effects group analyses.

The first set of fMRI analyses aimed at identifying regions that were commonly activated at the pre-training session for the criterion and transfer tasks. A conjunction approach (*7*) asserting that all contrasts involved were individually significant at the predefined threshold was used to investigate such overlap. All 22 participants were included in these analyses. For the *n*-back task, the analyses were restricted to 3-back, for which a behavioral training effect was observed. The conjunction analysis included [letter memory – baseline pre] and [3-back – baseline pre]. For the Stroop test, the conjunction analysis involved [letter memory – baseline pre] and [incongruent - congruent pre]. The analyses were thresholded at $p < .01$ after FDR correction for multiple comparisons (cluster size > 10 voxels). The Stroop data were also evaluated at a more liberal threshold ($p < .005$ uncorrected) to rule out the possibility that an absence of joint striatal activity for letter memory and Stroop was simply a thresholding effect.

Subsequent analyses focused on characterizing training-related activation changes [(task – baseline post) – (task – baseline pre)] for the criterion and transfer tasks at an uncorrected threshold of $p < .005$ (cluster size > 10 voxels). Regions that showed a training-related change were further analyzed relative to controls with 2 (Group: trained, control) x 2 (Session: pre-training, post-training) repeated measure ANOVAs. The ANOVAs were computed on data from the local maxima in regions showing pre-post changes for the training group. Regions for which a significant ($p < .05$) group x session interaction was found were further defined by BOLD bars. To qualify as a region showing a training-specific effect, the activity level for the training and control group had to be comparable at the first session, along with a selective or more pronounced change in activation level for the trained group at the second session.

On the basis of these initial analyses, subsequent analyses focused on common training-related activation changes for the letter memory and 3-back tasks in the striatum.

To investigate overlapping training-related changes for letter memory and 3-back, a conjunction analysis of between-session activation changes was conducted. Two contrasts were included in the conjunction: (1) [(letter memory – baseline post) - (letter memory – baseline pre)] and (2) [(3-back - baseline post) - (3-back - baseline pre)].

Finally, we performed a conjunction analysis with four contrasts to examine pre-post activity overlap for letter memory and 3-back in relation to initial pre-training activity: (1) [letter memory – baseline pre], (2) [(letter memory – baseline post) – (letter memory – baseline pre)], (3) [3-back - baseline pre] and (4) [(3-back - baseline post) – (3-back – baseline pre)].

Experiment 2.

Subjects

22 older adults were randomized to training or control groups. Participants were recruited through advertisements in one of the larger local newspapers. All participants who completed the study were paid SEK 1000 (approximately 160 USD). Three older persons declined to continue participation after the pre-training session. After exclusion of these individuals the participants ranged between 65 and 71 years of age (Table S1). The training and control groups did not differ significantly ($p > .05$) in age, years of education, depression (*1*), vocabulary (*2*), mental status (*3*) or mental speed measured by

pattern comparisons (*4*) and number copying (*4*). All participants reported being in good physical and mental health, not color blind, and right handed.

Both experiments were approved by the ethics committee of Umeå University. All participants were informed about the purpose of the investigation and their right to terminate participation at any point in time, and gave written informed consent to participate.

Procedure and tasks

The same procedure, training tasks, and fMRI data acquisition as in Experiment 1 were used here. The letter memory and *n*-back fMRI tasks from Experiment 1 were also included in Experiment 2.

Analyses of cognitive performance

The results from the pre- and post-training fMRI sessions were analyzed using 2 (Group: trained, control) x 2 (Session: pre-training, post-training) repeated measure ANOVAs. To compare the magnitude of gains, the effect size was calculated, as in Experiment 1, and differences between training and control groups were tested by two-tailed *t*-tests.

Analyses of brain activity

Functional images were analyzed with SPM2 and pre-processed as in Experiment 1. The first set of fMRI analyses aimed at identifying regions that were activated at the pre-training session for the criterion and transfer tasks for all 19 participants. These analyses were thresholded at $p < .001$ uncorrected (voxels > 10). To investigate if similar training-related changes and transfer effects were found as in Experiment 1, analyses with [(letter memory – baseline post) - (letter memory – baseline pre)] and [(3-Back –

baseline post) – (3-Back – baseline pre)] thresholded at $p < .005$ uncorrected (voxels > 10) were performed. To find regions for which a significant ($p < .05$) group x session interaction was found, ANOVAs were computed on data from the local maxima in regions showing pre-post changes for the training group. The same criteria to qualify as a region showing a training-specific effect were used as in Experiment 1. That is, the activity level for the training and control group had to be comparable at the first session, along with a selective or more pronounced change in activation level for the trained group at the second session.

In both Experiment 1 and 2, the peak coordinates from significant clusters are reported. Thus, two peaks may differ in one or several dimensions by a few mm, but still belong to the same cluster. Visualization of significant effects on brain templates and histogram plots were done using in-house developed software.

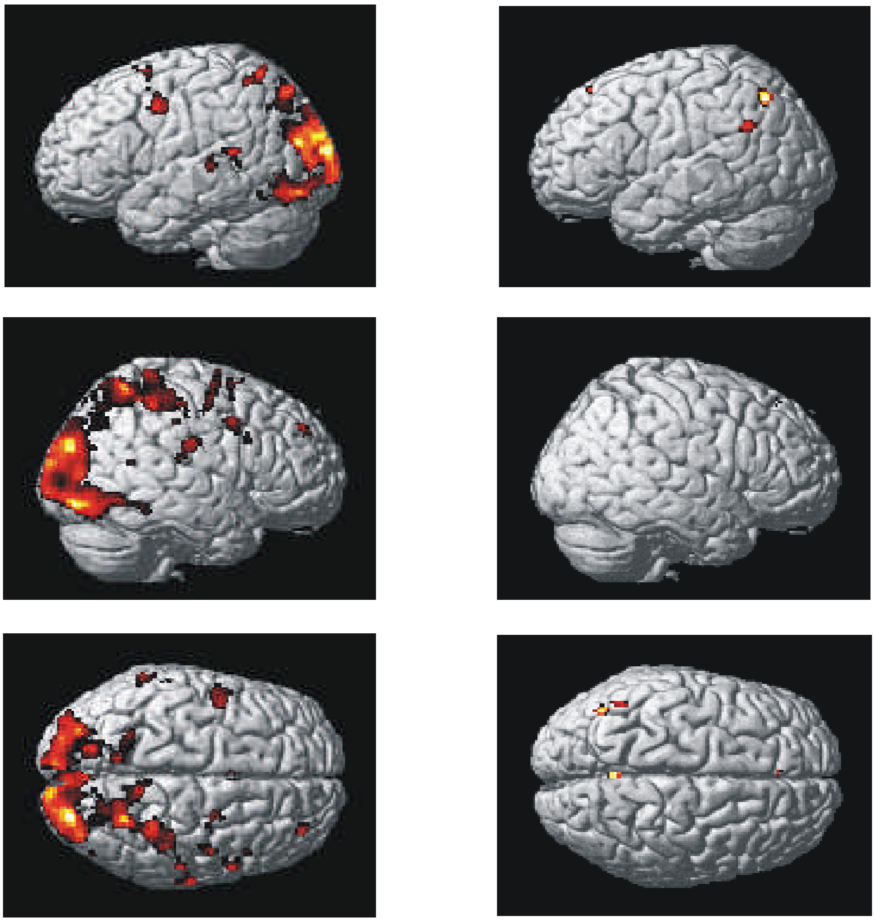

Fig. S1: Left column = (Letter Memory - Baseline) > (3-Back - Baseline); Right column = (3-Back - Baseline) > (Letter Memory - Baseline). Statistical parametric maps (n = 22) were projected onto cortical rendering templates, and were thresholded at p < .01 (FDR correction). Differential task activation was observed in the frontal, parietal, and occipital cortices. Letter memory was associated with greater activity in superior parietal cortex (BA 7; x, y, z = 28, -54, 62; -12, -60, 64), which, in conjunction with greater right dorsolateral frontal activity (BA 9; x, y, z = 34, 48, 38) may reflect higher demands on executive processing (*8*). 3-Back was associated with greater activity near left supramarginal gyrus (BA 39/40; x, y, z = -48, -54, 30; -42, -66, 48), which may relate to a greater need for comparison operations (*9*).

Table S1.
Subject Characteristics (M ± SD)

	Experiment 1		Experiment 2	
	Young-Training	Young-Control	Old-Training	Old-Control
Female/Male	*8/7*	*2/5*	*7/4*	*6/2*
Age	23.67 ± 2.92	23.43 ± 1.27	68.27 ± 1.79	68.38 ± 1.92
Years of education	13.40 ± 1.40	13.43 ± 0.73	14.55 ± 3.64	12.44 ± 2.58
BDI[a]	3.40 ± 3.27	5.00 ± 2.94	3.64 ± 3.33	6.50 ± 4.04
Vocabulary[b]	22.67 ± 2.61	20.86 ± 3.81	22.73 ± 4.82	23.88 ± 4.36
Mental status[c]	29.00 ± 0.93	29.29 ± 0.76	28.73 ± 0.91	28.75 ± 0.71
Pattern comparison[d]	21.27 ± 2.74	20.71 ± 5.02	15.50 ± 1.66	13.81 ± 2.55
Number copying[e]	50.87 ± 6.66	50.29 ± 9.30	44.77 ± 7.23	38.88 ± 7.47

[a]Beck Depression Inventory (BDI; Beck & Steer, 1996)
[b]SRB1 (Dureman & Sälde, 1959), maximum score = 30
[c]Mini Mental Test (Folstein et al., 1975)
[d]Pattern comparison (Salthouse & Babcock, 1991)
[e]Number copying (Salthouse & Babcock, 1991)

Table S2.
Pre- and Post-Training Performance in Criterion and Transfer Tasks (M ± SEM)

	Experiment 1		Experiment 2	
	Young-Training	Young-Control	Old-Training	Old-Control
Letter memory Pre	3.07 ± 0.52	3.00 ± 0.69	0.91 ± 0.37	0.25 ± 0.16
Letter memory Post	9.13 ± 0.24	4.71 ± 0.68	5.09 ± 0.91	0.63 ± 0.26
3-Back Pre Corr Yes Answer	26.53 ± 0.88	26.71 ± 2.14	19.36 ± 2.76	19.50 ± 2.78
3-Back Post Corr Yes Answer	32.13 ± 0.74	29.43 ± 2.02	21.82 ± 2.41	21.00 ± 2.37
3-Back Pre False Alarms	2.33 ± 0.86	1.29 ± 0.52	5.91 ± 2.65	2.38 ± 0.60
3-Back Post False Alarms	1.00 ± 0.39	0.86 ± 0.40	2.91 ± 0.46	3.13 ± 0.85
2-Back Pre Corr Yes Answer	32.67 ± 0.66	31.71 ± 1.55	22.18 ± 2.75	24.13 ± 2.68
2-Back Post Corr Yes Answer	34.20 ± 0.40	33.86 ± 0.91	27.55 ±1.50	26.38 ± 1.87
2-Back Pre False Alarms	1.67 ± 0.53	0.29 ± 0.18	5.45 ± 2.18	1.88 ± 0.64
2-Back Post False Alarms	0.87 ± 0.26	0.71 ± 0.29	2.00 ±0.57	2.13 ± 0.64
1-Back Pre Corr Yes Answer	33.67 ± 0.57	34.43 ± 0.81	27.09 ± 2.27	30.00 ± 1.13
1-Back Post Corr Yes Answer	34.67 ± 0.48	35.57 ± 0.30	31.91 ± 1.26	33.25 ± 0.86
1-Back Pre False Alarms	1.20 ± 0.33	0.57 ± 0.30	4.82 ± 2.59	1.38 ± 0.46
1-Back Post False Alarms	0.80 ± 0.30	0.14 ± 0.14	1.73 ± 0.54	1.50 ± 1.09
Stroop congruent Pre (RT)	741.36 ± 36.34	750.77 ± 32.99	-	-
Stroop congruent Post (RT)	675.82 ± 36.84	727.82 ± 36.22	-	-
Stroop incongruent Pre (RT)	871.15 ±39.04	914.20 ± 47.34	-	-
Stroop incongruent Post (RT)	775.69 ± 41.70	820.19 ± 32.28	-	-

Note. No Stroop data were available in experiment 2.

Table S3.

Training Related Increases and Decreases in Criterion and Transfer Tasks

	Area	Coordinates			Group * Session interaction
Experiment 1 - Younger adults					
Criterion task: letter memory					
Increases	L Striatum	-26	-4	-4	$F_{1.20} = 5.04$, $p < .05$
	R Striatum	28	-4	-2	$F_{1.20} = 9.40$, $p < .05$
	R Temporal lobe	26	6	-10	$F_{1.20} = 8.64$, $p < .05$
	R Occipital lobe	12	-78	16	$F_{1.20} = 8.31$, $p < .05$
Decreases	R Frontal lobe	36	60	4	$F_{1.20} = 7.05$, $p < .05$
	R Parietal lobe	32	-60	64	$F_{1.20} = 17.82$, $p < .001$
	R Parietal lobe	40	-54	62	$F_{1.20} = 11.39$, $p < .05$
Transfer task: 3-back					
Increases	L Frontal lobe	-32	36	4	$F_{1.20} = 9.10$, $p < .01$
	L Parietal lobe	-38	-22	66	$F_{1.20} = 7.40$, $p < .05$
	L Temporal lobe	-32	10	-10	$F_{1.20} = 7.52$, $p < .05$
	L Striatum	-30	4	6	$F_{1.20} = 6.57$, $p < .05$
	Brain stem	0	-38	-4	$F_{1.20} = 11.31$, $p < .01$
Experiment 2 - Older adults					
Criterion task: letter memory					
Increases	L Frontal lobe	-52	2	42	$F_{1.17} = 12.25$, $p < .005$
	L Parietal lobe	-54	-10	44	$F_{1.17} = 23.77$, $p < .001$
	R Parietal lobe	58	-4	32	$F_{1.17} = 8.62$, $p < .01$
	R Temporal lobe	46	14	-6	$F_{1.17} = 13.06$, $p < .005$
	L Cerebellum	-8	-78	-12	$F_{1.17} = 6.06$, $p < .05$
	L Cerebellum	-18	-60	-16	$F_{1.17} = 7.09$, $p < .05$
	L Striatum	-18	2	0	$F_{1.17} = 7.22$, $p < .05$
	L Striatum	-24	10	-2	$F_{1.17} = 7.42$, $p < .05$

Note: In Experiment 1, no training-related decreases were found for 3-back, and no training-related increases or decreases were found for 1-back, 2-back or Stroop. In Experiment 2, no training-related decreases were found in the criterion task and no increases or decreases were found in the transfer tasks.

Supplementary References

1. A. T. Beck, R. A. Steer. (Psykologiförlaget AB, Stockholm, 1996).

2. I. Dureman, H. Sälde. (Almqvist & Wiksell, Uppsala, 1959).

3. M. F. Folstein, S. E. Folstein, P. R. McHugh, *J. Psychiatr. Res.* **12**, 189 (1975).

4. T. A. Salthouse, R. L. Babcock, *Dev. Psychol.* **27**, 763 (1991).

5. A. Miyake *et al.*, *Cognit. Psychol.* **41**, 49 (2000).

6. J. H. Callicott *et al.*, *Cereb. Cortex* **9,** 20 (1999).

7. T. Nichols *et al.*, *NeuroImage* **25**, 653 (2005).

8. T. D. Wager, E. E. Smith, *Cognit. Affect. Behav. Neurosci.* **3**, 4 (2003).

9 J. D. Cohen *et al.*, *Nature* **386**, 6625 (1997).