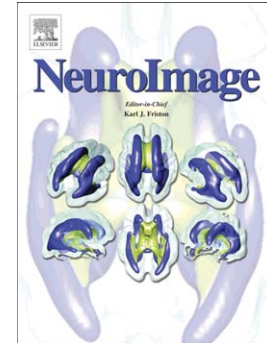# Accepted Manuscript

Neural correlates of training and transfer effects in working memory in older adults

Stephan Heinzel, Robert C. Lorenz, Patricia Pelz, Andreas Heinz, Henrik Walter, Norbert Kathmann, Michael A. Rapp, Christine Stelzel

Please cite this article as: Heinzel, Stephan, Lorenz, Robert C., Pelz, Patricia, Heinz, Andreas, Walter, Henrik, Kathmann, Norbert, Rapp, Michael A., Stelzel, Christine, Neural correlates of training and transfer effects in working memory in older adults, *NeuroImage* (2016), doi: 10.1016/j.neuroimage.2016.03.068

Neural correlates of training and transfer effects in working memory in older adults

Stephan Heinzel [1,2,3,4], *Robert C. Lorenz* [1,3,4,5], *Patricia Pelz* [4]; *Andreas Heinz* [3,4], *Henrik Walter* [3,4,6], *Norbert Kathmann* [1,3], *Michael A. Rapp* [# 2,4], *Christine Stelzel* [# 1,2,3,4,6,7]

[1]Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany;

[2]Social and Preventive Medicine, University of Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany;

[3]Berlin Center for Advanced Neuroimaging, Berlin, Germany;

[4]Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany;

[5]Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany;

[6]Berlin School of Mind and Brain, Germany.

[7]International Psychoanalytic University, Stromstr. 1, 10555 Berlin, Germany

*Corresponding author:

Dr. Stephan Heinzel, Dipl.-Psych., Humboldt-Universität zu Berlin, Wolfgang Köhler Haus, Rudower Chaussee 18, 12489 Berlin, Germany, Phone: +49 30 2093 - 4737;

Fax: +49 30 2093 – 4859, E-mail: stephan.heinzel@hu-berlin.de

[#] Both authors hold equal senior authorship.

1

Abstract

As indicated by previous research, aging is associated with a decline in working memory (WM) functioning, related to alterations in fronto-parietal neural activations. At the same time, previous studies showed that WM training in older adults may improve the performance in the trained task (training effect), and more importantly, also in untrained WM tasks (transfer effects). However, neural correlates of these transfer effects that would improve understanding of its underlying mechanisms, have not been shown in older participants as yet. In this study, we investigated blood-oxygen-level-dependent (BOLD) signal changes during n-back performance and an untrained delayed recognition (Sternberg) task following 12 sessions (45 minutes each) of adaptive n-back training in older adults. The Sternberg task used in this study allowed to test for neural training effects independent of specific task affordances of the trained task and to separate maintenance from updating processes. Thirty-two healthy older participants (60-75 years) were assigned either to an n-back training or a no-contact control group. Before (t1) and after (t2) training/waiting period, both the n-back task and the Sternberg task were conducted while BOLD signal was measured using functional Magnetic Resonance Imaging (fMRI) in all participants. In addition, neuropsychological tests were performed outside the scanner. WM performance improved with training and behavioral transfer to tests measuring executive functions, processing speed, and fluid intelligence was found. In the training group, BOLD signal in right lateral middle frontal gyrus/ caudal superior frontal sulcus (Brodmann area, BA 6/8) decreased in both the trained n-back and the updating condition of the untrained Sternberg task at t2, compared to the control group. FMRI findings indicate a training-related increase in processing efficiency of WM networks, potentially related to the process of WM updating. Performance gains in untrained tasks suggest that transfer to other cognitive tasks remains possible in aging.

2

Introduction

Age-related reduction in working memory (WM) performance has been related to changes in brain functioning (for reviews, see Eyler et al., 2011; Nyberg et al., 2012; Rajah and D'Esposito, 2005). Studies comparing neural activity during verbal WM tasks between older and younger participants have found that younger adults asymmetrically recruited areas in left fronto-parietal areas, while older adults showed greater activity also in homologous regions in the right hemisphere (Cabeza et al., 2004, 2002; Reuter-Lorenz et al., 2000). More recent studies (Heinzel et al., 2014a; Nagel et al., 2011) have indicated that age-associated activation differences depend largely on task difficulty. Compared to younger adults, older adults showed similar performance but greater activation in WM-related brain areas at low WM load but reduced performance and lesser activation at high WM load in a recent sample of younger and older adults that included a subsample of the current study (Heinzel et al., 2014a). These findings can be explained by the notion of reduced neural efficiency and capacity in older adults (Barulli and Stern, 2013) and the compensation-related utilization of neural circuits hypothesis (CRUNCH, Reuter-Lorenz and Cappell, 2008). The CRUNCH model suggests that older adults recruit more neural resources to achieve a similar performance as younger adults at relatively low task demands (reduced processing efficiency in older adults). It has been argued that older adults utilize cognitive control strategies already at low task difficulty levels to compensate for structural and functional decline (Grady, 2012). However, according to the CRUNCH model, attempted compensation fails at high task difficulty because older adults are unable to further exceed their neural activation level and performance collapses due to a reduced capacity in older adults (Nyberg et al., 2009;

3

Schneider-Garces et al., 2010). Cognitive training is thought to enable older adults to perform low and medium levels of WM tasks below their capacity limit due to the development of more efficient processing strategies. Therefore, WM training is expected to lead to activation decreases only at low and medium difficulty, but not at high difficulty levels (Lustig et al., 2009).

Cognitive training research has shown that WM training might have the potential to slow down or even restore some aspects of age-related decline in WM functioning (e.g. (Brehmer et al., 2011; Dahlin et al., 2008; Heinzel et al., 2014a, 2014c; Li et al., 2008; Richmond et al., 2011). While several studies have investigated neural correlates of WM training gains in younger, there are only few studies in older adults (Buschkuehl et al., 2012; Klingberg, 2010). Erickson et al., (2007) found that dual-task performance gains following training were related to decreased activity mainly in right dorsolateral prefrontal cortex (DLPFC) in older adults. Similarly, (Brehmer et al., 2011), reported an activity decrease in WM-associated areas in right fronto-parietal regions in response to behavioral training gains in a delayed recognition task. If training enables older participants to increase the efficiency of their WM processing as suggested by the CRUNCH model (Reuter-Lorenz and Cappell, 2008), activity is expected to decrease particularly at 1-back (low difficulty) and 2-back (medium difficulty) following training. Taken together, behavioral training gains seem to be associated with changes in neural activity during the performance of the trained task. However, results might be relatively task-specific since neural correlates of transfer to other WM tasks have only been reported in one study in younger but not in older adults as yet (Dahlin et al., 2008). Furthermore, it is not known which components of WM are associated with training-related activity changes.

According to the WM model of Baddeley (2000), WM refers to "a limited capacity system allowing the temporary storage and manipulation of information" (Baddeley, 2000,

4

p.418). Although there has been a proliferation of cognitive psychological theories on WM since Baddeley and Hitch's (1974) model (for a comparison of models, see e.g. Chein and Fiez, 2010), most WM models suggest a distinction between at least two components of WM, namely maintenance and executive control of information. Both components are thought to be involved in the performance of the n-back WM task. While n-back training was found to be effective in older adults (Heinzel et al., 2014a, 2014c; Li et al., 2008), as yet, it is unclear whether training gains in the n-back task refer to an improved ability to maintain or to update information or both. If activity changes related to training gains within a cognitive process (e.g. updating) are not just task-specific, common activity changes in similar tasks could be expected (Buschkuehl et al., 2012; Dahlin et al., 2008; Gray et al., 2003; Jonides, 2004). Thus, training-related changes in brain activity during n-back performance might overlap with activity changes in the maintenance or the updating component of WM in an untrained task depending on the specific process component mediating training gains.

There has been a debate whether subcomponents of WM such as updating or maintenance may be associated with specific brain areas. Modality-specific activity patterns in ventral vs. dorsal posterior lateral frontal cortex and more posterior brain regions have been shown consistently for the maintenance of verbal vs. spatial information (Curtis and D'Esposito, 2003). In line with current conceptions of PFC function (Frank et al., 2001; Fuster, 2004; Miller and Cohen, 2001), WM updating can be understood as a highly interactive executive control process, not just involving DLPFC, which has been most prominently associated with executive control (Baddeley, 2003; Collette and Van der Linden, 2002; D'Esposito et al., 2000, 1995; Mohr et al., 2006). Instead, WM updating seems to rely on a distributed network also including ventrolateral PFC and areas of the dorsal attention system located in lateral premotor cortex (LPMC)/caudal superior frontal sulcus (cSFS), in posterior parietal cortex (PPC, for recent reviews, see Linden, 2007; Nee et al., 2013) and

5

subcortical regions (Frank et al., 2001), which partly overlap with regions related to maintenance. Therefore, activity changes within these areas after n-back training might be either specific to maintenance or updating or common to both.

Finally, an important question in cognitive training research concerns the degree to which training effects are transferable to tasks in other cognitive domains ("far transfer", Klingberg, 2010; Lustig et al., 2009; Noack et al., 2009; Zelinski, 2009). In the past decade, WM has gained attention in transfer research, presumably because WM is believed to be a central mental capacity that has been shown to be closely linked to other cognitive domains, such as executive functions (e.g. Chen and Li, 2007; Conway et al., 2003), processing speed (e.g. Burgaleta and Colom, 2008; Clay et al., 2009; Salthouse, 1996), and fluid intelligence (e.g. Ackerman et al., 2005; Kyllonen and Christal, 1990).

Since n-back is considered to be an executively demanding WM task (Veltman et al., 2003), we expected training gains in n-back to be related to the executive component of WM and thus find overlapping activity changes between n-back and the updating component of an untrained WM task. Furthermore, we expected to detect behavioral transfer to tests of executive functions and short-term memory outside the MRI scanner. By including additional executive neuropsychological tests (such as the Stroop word/color interference test, Stroop, 1935), we will investigate if transfer effects are specific to updating or if they are also found in other domains of executive functions (such as "inhibition" according to the model by Miyake et al., 2000). To test these hypotheses in the present study, a sample of healthy older adults was assigned to either an n-back training group or to a no-contact control group. Before and after training/waiting period, a parametric n-back task was performed during fMRI measurement and a set of cognitive tasks was administered. An additional WM paradigm (Sternberg delayed match to sample task) for which WM component processes (i.e. maintenance vs. updating) could separately be analyzed was administered before and after

6

training to investigate common activity changes between n-back and Sternberg task. That way, it could be tested whether the neural correlates of transfer effects in WM are associated with maintenance or updating processes.

Taken together, the specific aims of this study were to (i) investigate whether training-induced changes in neural activation in older adults reflect increases in processing efficiency as postulated by the CRUNCH model; (ii) test whether these neuronal activation changes vary with WM load as predicted by task demand to activation functions. Regarding transfer effects, we aimed to investigate (iii) which components (updating versus maintenance) of the working memory system are involved in transfer effects, (iv) whether transfer effects go along with neuronal activation overlap, and (v) whether such overlap is related to near and far behavioral transfer effects.

Methods

Participants

Sixteen control participants were carefully matched one by one to sixteen training participants according to their age (+/- 2 years), sex, and education level (+/- 3 years of education) to assure parallelization of both groups. The first half of the control group was recruited after the first half of the training group was included, the second half of the control group after testing the second half of the training group. The age range was 60 to 75 years. All participants were recruited via newspaper announcements in Berlin, Germany. All were native German speakers, right-handed, had normal or corrected-to-normal vision, no history of any neurological or psychiatric diseases, and did not take any psychiatric medication. Mini Mental Status Examination (MMSE, Folstein, Folstein, & McHugh, 1975) was 27 or above in all participants. All participants were suitable for functional Magnetic Resonance Imaging (fMRI) and none of the participants took any anti-hypertensive medication or had a thyroidal

7

dysfunction, which could have influenced the blood-oxygen-level-dependent (BOLD) signal. One participant of the training group had to be excluded due to a technical failure during the fMRI scanning and two participants in the control group had to be excluded due to profound signal loss in ventral regions in prefrontal, temporal, and occipital cortex, as well as subcortical regions most likely due to numerous dental implants. Therefore, the final sample consisted of 15 participants (9 women) in the training group (mean age = 66.07 [± 4.04] years, education = 15.43 [± 3.34] years, MMSE = 29.13 [± 1.25] points) and 14 participants (9 women) in the control group (mean age = 66.00 [± 4.82] years, education = 16.11 [± 2.99] years, MMSE = 29.43 [± 0.76] points). The groups did not differ significantly (p> .45) in age, sex distribution, education level and MMSE. Please note that results from a subsample have previously been reported in a region of interest (ROI)-based analysis (Heinzel et al., 2014a). The study was approved by the ethics committee of the Charité - Universitätsmedizin, Berlin. Written informed consent was obtained from all participants prior to investigation. Participants who completed the study received a monetary reimbursement.

Design and procedure

At pre-test (t1), all participants completed a set of neuropsychological tests and participated in two separate fMRI sessions (first n-back and then Sternberg task). Subsequently, the training group participated in a WM training program for 4 weeks (12 sessions), whereas the control group did not receive any intervention. Note that unspecific effects of the intervention should affect all WM component processes and thus should not lead to differential training and transfer effects. After the training/waiting period (post-test, t2), all participants returned for two separate fMRI sessions (first n-back and then Sternberg task). Finally, all participants completed the same set of neuropsychological tests from t1 again at t2.

Materials and tests

For neuropsychological screening and the examination of possible transfer effects of WM training to other cognitive tasks, the following tests were conducted at t1 and t2: Digit Span Forward and Backward (Wechsler, 1987) were included as proxies for short-term memory, Digit Symbol Substitution (Wechsler, 1987) and D2 (Brickenkamp, 2002) for processing speed, Stroop Word/Color Interference (Stroop, 1935) and Verbal Fluency (German version of the Controlled Oral Word Association Test (COWAT, Benton and Hamsher, 1989)) for executive functions, and Raven's Standard Progressive Matrices (Raven's SPM, Raven et al., 1990) and the LPS Figural Relations subtest of a German intelligence test (Leistungspruefsystem LPS, Horn, 1983) for fluid intelligence. Parallel versions were used for Stroop, Raven's SPM, and LPS Figural Relations to avoid memory effects at t2. In Digit Span Forward and Backward, two trials of each list length were presented. If participants failed to repeat both trials of a certain list length, the assessment of this task was terminated. The score used in the following analyses was determined by the length of the longest correctly repeated list. In the D2-test, participants were instructed to cross out the letter 'd' with exactly two dashes, but neither the letter 'p' nor the 'd' with a different number of dashes than two. The score used in this study represents the amount of crossed out 'd's within 4min and 40 sec. In Digit Symbol, participants were asked to copy symbols as quickly as possible into empty boxes located below a random sequence of numbers ranging from 1 to 9 according to a specific coding key. The score used for analyses was the number of correct symbols completed within 60 seconds. In the Verbal Fluency task, participants were asked to generate as many words as possible starting with the letter 'S' within 60 seconds (not including proper names or names of places and cities). In the Stroop task, participants were first instructed to name the color of words (task A), then to name the

9

meaning of words (task B), and finally to name the color of words while inhibiting the meaning of the words (a different color, task C). The outcome score in each of the three tasks represents the time required to finish each task. The Stroop Interference score in this study was calculated using the following equation (Golden, 1978): Stroop Interference = task C − [(task A * task B) / (task A + task C)]. To solve the fluid intelligence tasks (Raven's SPM and LPS Figural Relations), participants were required to identify patterns of nonverbal symbols: In Raven's SPM, they were instructed to find a matching item to complete a pattern, while in the Figural relations, they had to mark the non-matching item of a pattern of symbols. Both reasoning tasks were timed and the scores were derived from the number of correct items accomplished within 7.5 minutes (Raven's SPM) or 3 minutes (LPS Figural Relations), respectively. To rule out clinically relevant mild cognitive impairment (Petersen et al., 1999), we used delayed recall subtest from the neuropsychological test battery of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD, Morris et al., 1989) and the Mini-Mental State Examination (MMSE, Folstein et al., 1975), a short screening test addressing basic cognitive functions such as orientation, memory, simple computational skills, reading and visuo-constructive abilities. Data was checked by a dementia expert at the clinic and participants were excluded if they scored below 27 points on the MMSE or achieved below 1.5 standard deviations of age, education, and gender matched norms in the CERAD delayed recall test.

N-back paradigm during fMRI

A computerized version of the n-back paradigm with numerical stimuli (Cohen et al., 1997) was used in this study (see Figure 1, panel A). The n-back task consisted of 2 runs. In each run, 16 blocks were presented in 4 different pseudo-randomized orders counterbalanced across subjects. The block lengths were 16 seconds (blocks with an inter trial interval [ITI] of

500ms) or 32 seconds (blocks with an ITI of 1500ms). The total duration of the task was 22 minutes. The n-back task was presented using Presentation software (version 14.9, Neurobehavioral Systems Inc., Albany, CA, USA). WM load (0-, 1-, 2-, and 3-back) was varied between blocks. In each block, 16 randomly created digits from 0 to 9 were presented in the center of a black screen one at a time for 500ms; the occurrence of 5 target stimuli in each block was pseudo-randomized. Targets were defined as re-occurrence of a number previously presented 1, 2, or 3 trials before (1-, 2-, or 3-back condition, respectively). In the 0-back condition, the target was defined as the number '0'. Response time was from the onset of a target until the onset of the next stimulus. Each condition (defined by WM load and ITI) was repeated 4 times. Participants were instructed to press a button with their right thumb when they recognized a target. After each block, a white fixation cross was presented in the center of a black screen for 12 seconds.

-----------------------------Please insert Figure 1 approximately here-------------------------------

Sternberg paradigm during fMRI

The delayed match-to-sample ("Sternberg") task (see Figure 1, panel B) was administered in a 2 x 2 design where WM load (3 vs. 5 letters) and updating demands (maintenance only vs. maintenance plus updating) were manipulated orthogonally. The task consisted of three phases: encoding, delay (separated into an early delay period and a task delay period), and probe phase. In the encoding phase, either three or five sequentially presented letters had to be encoded. In the early delay phase, participants maintained the encoded set of letters in memory. Before the task delay phase, a written cue indicated which task to perform on the letters. In the maintenance condition (cued by the word 'maintain', in German: 'merke'), participants continued to maintain the letters in the order of presentation (upper stream in Figure 1B). In the updating condition (cued by the word 'substitute', in

11

German: 'ersetze') either 1, 2, or 3 new letters were presented pseudo-randomly. This was done to make sure participants would update one item at a time; for analysis purpose, the amount of updated items was pooled, but see Supplementary Table S1b for the performance data dependent on the number of items during the delay period. Updating involved mentally deleting the respective number of letters from the maintained sequence, starting with the first item from the currently maintained list and adding the new item(s) to the end of the list. Thus, updating 'z k r' with 't' resulted in 'k r t' (lower stream in Figure 1, panel B) - so the number of letters to be maintained (WM load) remained constant at either 3 or 5 letters, but the specific contents had to be updated. In the maintenance condition, participants saw hash keys (#) serving as placeholders in the presentation to allow for comparison with the updating condition. In the probe phase, a probe stimulus, consisting of a capital letter required retrieval of information from WM. Timing information is given in Figure 1, panel B. With their left and right thumbs, participants indicated via button press whether the respective letter was part of the most recent memory set (right thumb) or not (left thumb). Together, this resulted in four task conditions: maintain-3, maintain-5, update-3, update-5. Four experimental runs were presented, each run comprising 24 trials, resulting in a duration of ~45 minutes for the whole experiment. Altogether, there were 24 trials for each of the four conditions. Conditions were evenly distributed across runs and restrictions during randomization ensured that transitions between conditions were equal.

N-back training procedure

The training group participated in an n-back training program over a period of 4 weeks with 3 sessions per week, resulting in 12 training sessions. Participants accomplished 3 runs of the n-back task in each training session, lasting about 45 minutes. Each run consisted of 12 blocks. At run 1 in session 1, all participants began the training with the difficulty level 1 (4

blocks of 0-back, 4-blocks of 1-back, and 4-blocks of 2-back, at an ISI of 1500ms). Adaptive practice (Doumas et al., 2009) was used throughout all 12 training sessions in order to keep the task demanding. The difficulty level of the task varied across training runs according to individual performance. Task difficulty was increased by introducing higher WM load levels and by shortening the ISI (Heinzel et al., 2014a, 2014b, 2014c). If a participant successfully completed one run with a hit rate of 80% or above within each block and with a false alarm rate below 15%, the next difficulty level was introduced in the following run. From level 1 to level 3, ISI was gradually decreased from 1500 to 500ms in steps of 500ms. At level 4, the next n-level was introduced (3-back), and 0-back was removed, i.e. participants completed 1-, 2-, and 3-back tasks. In addition, ISI was set back to 1500ms. At level 7, 4-back was introduced and 1-back was removed.

FMRI data acquisition

In the beginning of each scanning procedure, one T1-weighted 3D fast spoiled-gradient echo pulse sequence was obtained (TR=7.8ms, TE=3.164ms, flip angle=20°, FoV=256x256 mm², matrix size= 256x256, 176 sagittal slices with 1mm thickness, voxel size=1x1x1 mm³). Functional data was obtained using a gradient echo-planar imaging (EPI) sequence with standard parameters (TR= 2000ms, TE_{GE} = 35ms, TE_{Trio}= 32ms, flip angle= 80°, matrix size= 64x64, voxel size= 3.1x3.1x3.8mm). Thirty-one slices were acquired approximately axial to the bicommissural plane. Whole-brain fMRI data was collected at two different scanners due to a forced change of the scanner in the course of the study. 12 (12 in training group) participants were scanned at Charité Virchow Campus, Berlin with a 3 Tesla GE Signa Excite (General Electrics Medical Systems, Milwaukee, USA) and 17 (3 in training group) participants at Charité Campus Mitte, Berlin with a 3 Tesla Magnetom Trio Tim MR

13

system (Siemens, Erlangen, Germany). Scanner site was included as covariate in all second level statistical analyses.

Statistical analyses of behavioral data

Statistical analyses of the behavioral data were performed using SPSS version 21.0 (SPSS, Chicago, IL). Kolmogorov-Smirnov test for deviation from a normal distribution indicated that all analyzed variables met assumptions for parametric testing. For the n-back task, the comparison between groups and test intervals were then conducted using a repeated measures general linear model (GLM) analysis of variance with the between-subjects factor group (training vs. control group) and the within-subject factors time (t1 vs. t2) and WM load (0-, 1-, 2-, and 3-back). For the Sternberg task, repeated measures GLMs for the performance data were conducted with the between-subject factor group (training vs. control group) and the within-subject factors time (t1 vs. t2), updating demands (maintenance vs. updating) and load (3 vs. 5 letters). Changes in the neuropsychological measures outside the scanner were analyzed by a multivariate repeated measures GLM including all 8 neuropsychological tests and follow-up univariate repeated measures GLMs for each test separately with the between-subjects factor group and the within-subject factor time.

MR image processing and analysis

All fMRI analyses were carried out with SPM8 (http://www.fil.ion.ucl.ac.uk/spm). First, origins in the anatomical and functional images were set manually to correspond approximately to the anterior commissure. To improve data quality, the slice repair function of the artrepair toolbox (http://www.nitrc.org/projects/art_repair/) was applied. Each participant's functional data set was slice-time corrected (Sternberg task only due to its event related nature) and head motion corrected. The anatomical image was then coregistered to the

14

mean functional image per subject. After segmentation of the anatomical image, the functional data were spatially normalized into the standard MNI atlas space using the normalization parameters from the anatomical segmentation. Data were then smoothed with an 8-mm full width at half maximum Gaussian kernel and high-pass filtered during statistical analysis (128s). We applied the general linear model (GLM) for serially auto-correlated data (Friston et al., 1994) to determine task-related activity, applying a canonical hemodynamic response function. Runs were modeled as separate sessions.

FMRI data analyses n-back task

All experimental conditions (blocks of 0-, 1-, 2-, and 3-back) were included in the GLMs as separate boxcar regressors ("block design"). The cues, responses, and realignment parameters from the motion correction were included as 8 regressors of no interest in the model. On the subject level, linear contrast images were calculated for 0-, 1-, 2-, and 3-back blocks separately for each time point (t1 and t2) and contrasted against implicit baseline (fixation cross). On the group level, these contrasts were entered into separate flexible factorial models for each WM load level (0-, 1-, 2-, and 3-back), with the factors subject, group (training vs. control group) and time (t1 vs. t2), and the covariate "site". A tissue probability map included in SPM 12 (derived from the IXI data set, http://www.brain-development.org), thresholded at $p = .3$ for gray matter, was used as an explicit mask in all fMRI analyses to ensure that all reported activations are likely to be located in gray matter. Main effects of task are reported at $p < .05$, FWE-corrected for the whole brain ($k > 10$ voxels). Whole brain analyses of the group by time interaction effects were thresholded at $p < .05$, alphasim-corrected (alphasim REST toolbox implementation, http://www.resting-fmri.sourceforge.net). We used a Monte-Carlo simulation correction (10000 iterations) with a voxelwise threshold of $p < .005$. Only clusters above a cluster size that yielded an alphasim

15

correction threshold of p < .05, corrected, were described in the results section and in Table 3. The main focus of the current study was the investigation of neural correlates of operation-specific (maintenance vs. updating) transfer effects. Therefore, we increased the power for these analyses by not including WM load as an additional factor when investigating neural overlap with activation changes in the Sternberg task. Since training-related activation changes were expected in 1- and 2-back, a combined contrast was computed for 1- & 2-back when building a mask image for the analyses of neural correlates of transfer to the Sternberg task.

FMRI data analyses Sternberg task

The GLMs included separate regressors (event-related design) for all experimental conditions distinguishing the four task phases, i.e., encoding, early delay, task delay, and probe. In addition, the model included regressors for incorrectly answered trials and the six realignment parameters derived from preprocessing. To separate activity related to maintenance vs. updating processes, analyses were restricted to the task delay phase. On the subject level, linear contrasts were defined (regressors of interest contrasted against implicit baseline [fixation cross]) that were then entered into flexible factorial models on the group level with factors subject, group (training vs. control group), and time (t1 and t2), and the covariate "site" separately for each task contrast of interest. As in the n-back analyses, a tissue probability map (.3 threshold for gray matter) was used as an explicit mask in all fMRI analyses. To assess the task-related networks, task contrasts were defined at t1 per task cell, averaged across both groups and thresholded at p < .05, FWE-corrected for the whole brain (k > 10 voxels). Likewise, differential contrasts per operation (updating > maintenance and load-5 > load-3) were calculated for t1, across groups to determine the relevant networks per operation before training. To investigate training-specific changes in brain activity, the group

16

by time interaction effects were calculated per operation. Comparable to the n-back fMRI analyses, the two different WM load levels (3 and 5 letters) of the Sternberg task were averaged and whole brain analyses of the group by time interaction effects were thresholded at p < .05, alphasim-corrected using a Monte-Carlo simulation correction with 10000 iterations and a voxelwise threshold of p < .005. Again, only clusters above a cluster size that yielded an alphasim correction threshold of p < .05, corrected, are described in the results section and in Table 3.

Analysis of neural overlap between n-back and Sternberg tasks

A neural overlap was defined as commonly activated voxels in both n-back and Sternberg tasks that were found in significantly activated clusters in each individual task and that did not show baseline differences at t1. The overlap cluster is depicted in Figure 3. We used the MarsBaR toolbox (http://www. marsbar.sourceforge.net) to compute percent signal change and according effect sizes in the overlap peak voxel (peak of the Sternberg group by time interaction masked by the n-back group by time interaction) for all conditions in the n-back and Sternberg tasks (see Table 4).

<div align="center">

Results

</div>

Behavioral results

N-back performance and training gains

The groups did not differ in n-back performance at t1 (all p's> .18). To test whether improvements in n-back performance (% correct, defined as hit rate minus false alarm rate) can be related to the training, and whether training gains differ between different WM load levels, a group by time by WM load repeated measures GLM was conducted and revealed a significant 3-way interaction (F(3,81) = 5.13, p = .003, partial η² = .16). See Table 1, panel A

17

and Figure 1, panel A for means and standard errors per condition. The main effect of time ($F(1,27) = 41.67$, $p < .001$, partial $\eta^2 = .61$) indicated general performance improvement over time. The main effect of WM load ($F(3,81) = 302.99$, $p < .001$, partial $\eta^2 = .92$) showed that performance differed between WM load levels. No main effect of group was found ($F(1,27) = 2.15$, $p = .154$, partial $\eta^2 = .07$). The group by time interaction ($F(1,27) = 24.07$, $p < .001$, partial $\eta^2 = .47$) indicates that the training group showed generally stronger improvement in n-back performance compared to the control group. The WM load by time interaction ($F(3,81) = 8.10$, $p < .001$, partial $\eta^2 = .23$) shows that the magnitude of improvement over time differed between the WM load levels. No significant group by WM load interaction was found ($F(3,81) = 1.96$, $p = .126$, partial $\eta^2 = .07$). Follow-up t-tests showed significant performance improvements over time in the training group in 1-back ($t(14) = 3.67$, $p = .003$, $d = .97$), 2-back ($t(14) = 7.25$, $p < .001$, $d = 1.86$), and 3-back ($t(14) = 4.08$, $p = .001$, $d = 1.19$). Except for one participant who showed no training gain, all individuals in the training group improved their n-back performance from t1 to t2. No significant performance changes in any of the WM load levels were found in the control group (all p's > .10, d's < .50).

Analyses of reaction time (RT) data for the n-back task are reported in the Supplementary results and in Supplementary Table S1a, Panel A.

-----------------------------Please insert Table 1 approximately here-----------------------------

Behavioral transfer to the Sternberg task

In the Sternberg task, performance (% correct, defined as % hits) was analyzed with a repeated measures GLM including the between subject factor group (training vs. control) and within subjects factors updating demands (maintenance vs. updating), WM load (3 vs. 5) and time (t1 vs. t2). See panel B of Table 1 and panel B of Figure 1 for means standard errors and

18

effect sizes per condition. In addition, performance data for the combined update-3 and update-5 condition are shown in Table 1, as this combined condition will be further analyzed in the fMRI data. The 4-way interaction was not significant, also no interaction effects with the factors group or time were found in the full model. However, the analysis revealed generally lower performance in the updating compared to the maintenance condition, $F(1, 27) = 44.18$, $p < .001$, partial $\eta^2 = .62$ and lower performance at WM load of 5 compared to 3, $F(1, 27) = 60.28$, $p < .001$, partial $\eta^2 = .69$. A significant interaction of the factors updating demands and WM load, $F(1, 27) = 8.63$, $p = .007$, partial $\eta^2 = .24$, reflects that the reduction in performance from WM load 3 to WM load 5 was stronger in the updating condition than in the maintenance condition. The group by time interactions separately for each condition as well as effect sizes of follow-up T-tests are shown in Table 1, panel B. These post-hoc analyses indicate significant improvements in maintain-5 ($t(27) = 2.29$, $p = .038$) and the combined update-3 & -update-5 condition ($t(27) = 2.38$, $p = .032$) after n-back training in the training group that were not found in the control group (all p's > .35, see Table 1, panel B for all d's). Additional post-hoc t-tests confirmed that there were no significant differences in the performance between the training group and the control group at t1 in any of the four conditions and the combined update-3 & update-5 condition (all p's > .14, see Table 1, panel B for all d's). Note, that load conditions were pooled for fMRI analyses to increase the power for the consideration of operation-specific (maintenance vs. updating) effects. Reaction time patterns are reported in the Supplementary Table S1a, panel B.

Transfer to neuropsychological tests outside the MRI scanner

Training and control group showed no significant differences in any of the neuropsychological tests at t1 (all p's > .14). To investigate transfer to neuropsychological tests outside the MRI scanner, a group by time multivariate GLM was conducted including all

19

8 neuropsychological tests (Digit Span Forward, Digit Span Backward, D2 Test, Digit Symbol Substitution, Verbal Fluency, Stroop Interference, Raven's Standard Progressive Matrices, and LPS Figural Relation Test). A significant group by time interaction was found ($F_{(8, 20)}$ = 3.54, p = .010, partial η^2 = .59), showing larger improvements in the neuropsychological tests in the training compared to the control group. A significant main effect of time ($F_{(8, 20)}$ = 17.03, p < .001, partial η^2 = .87) indicates general improvements from t1 to t2 assessments. Follow-up group by time GLMs were performed for each test separately. A transfer effect was defined as a significant group by time interaction, where the training group increases more than the control group in the performance of a given test. Results and statistical values are shown in Table 2. Transfer effects were found in D2 ($F_{(1,27)}$= 10.32, p= .003, partial η^2= .28), Stroop Word/Color Interference ($F_{(1,27)}$= 6.20, p= .019, partial η^2= .19), and LPS Figural Relations ($F_{(1,27)}$= 4.53, p= .043, partial η^2= .14). The interaction effects reflect that the task performance in the training group increased while task performance did not increase significantly in the control group. Main effects of time were found in Digit Span Backward ($F_{(1, 27)}$ = 5.36, p = .028, partial η^2= .17), D2 Test ($F_{(1, 27)}$ = 8.58, p = .007, partial η^2= .24), Digit Symbol Substitution ($F_{(1, 27)}$ = 14.69, p = .001, partial η^2= .35), Stroop Interference ($F_{(1, 27)}$ = 8.40, p = .007, partial η^2= .24), and LPS Figural Relation Test ($F_{(1, 27)}$ = 13.30, p = .001, partial η^2= .33).

-------------------------------Please insert Table 2 approximately here-------------------------------

FMRI results

Main effects of the n-back task at t1

The n-back task main effects (0-, 1-, 2-, and 3-back) were tested against the implicit baseline across both groups at t1 and revealed a pattern of activity mainly in fronto-parietal areas that has been frequently reported in n-back studies (for review, see Owen et al., 2005).

20

Comparing these task main effects between training and control group at t1 showed no significant differences (thresholded at p < .05 FWE-corrected for whole brain, k > 10 voxels). As displayed in the upper panel of Figure 2, main foci of the n-back task activations were found in the lateral middle and superior frontal gyri including dorsolateral prefrontal and premotor cortex in both hemispheres. The cluster extended ventrally to inferior frontal gyrus and insula, caudally to the precentral gyrus, and medially to the supplementary motor area and the anterior cingulate gyrus. N-back-related activations were also found in inferior and superior parietal lobules, as well as in visual regions in the occipital cortex. Subcortically, n-back was associated with striatal activity and activity in the thalamus and the cerebellum. Anatomical locations and MNI coordinates for all n-back conditions are shown in Supplementary Table S2.

-------------------------------Please insert Figure 2 approximately here-------------------------------

Main effects of the Sternberg task at t1 and overlap with n-back main effect

First, effects of the different task conditions during the task phase were tested against baseline across both groups for t1 (see middle panel of Figure 2 and Supplementary Table S3), revealing the involvement of well-known fronto-parietal WM regions in the maintain-5, update-3 and update-5 conditions and to a lower degree for maintain-3. Greater demands during updating compared to maintenance (main effect updating demands) were reflected in higher activity in bilateral fronto-parietal regions, while higher WM load (main effect load) was associated with activity in anterior portions of lateral PFC, bilaterally (see lower panel of Figure 2). As shown in Figure 2 and Supplementary Tables S2 and S3, the Sternberg updating task involved fronto-parietal areas that largely overlap with areas activated by the n-back task. Comparing these task main effects between training and control group at t1 showed no significant differences in the fronto-parietal WM network at t1 (p < .05 FWE-corrected for

21

whole brain, k > 10 voxels). The only difference between groups was greater activity in primary visual cortex for the maintain-3, and maintain-5 conditions for the control group compared to the training group.

Neural correlates of training effects in the n-back task

Training-specific activity changes were calculated for 0-, 1-, 2, and 3-back separately, and for the combined 1- & 2-back contrast by performing flexible factorial models with factors subject, group (training group vs. control group), and time (t1 vs. t2). Scanner site was included as a covariate. Results are reported at a significance threshold of p < .05 alphasim cluster- corrected in Table 3 and depicted in Figure 3, panel A.

In the combined 1- & 2-back contrast (alphasim cluster threshold = 91 voxels), stronger activity decreases in the training group compared to the control group at t2 were found in a medial cluster including medial frontal gyrus/ anterior cingulate gyrus, extending caudally to mid-cingulate gyrus, and supplementary motor area in both hemispheres and a right lateral cluster in the middle and superior frontal gyrus including dorsolateral prefrontal and premotor cortex. A third cluster showed training-related decreases in the right supramarginal gyrus, inferior parietal lobule, and angular gyrus. As indicated by pre-post differences in PSC (see Figure 3 panel C), activity in the training group decreased from t1 to t2, while no activity changes were observed in the control group. As shown in Figure 3 panel A, a similar activation pattern as in the combined 1- & 2-back contrast was found when testing the 2-back condition separately. In the separate 2-back condition (alphasim cluster threshold = 86 voxels), the medial frontal cluster extended further to mid-cingulate gyrus and medial parts of paracentral lobule and laterally to superior frontal gyrus in the right hemisphere. A second cluster included medial parts of mid-cingulate and paracentral gyrus also in the left hemisphere. A third left hemispheric cluster included superior and middle

temporal gyrus as well as angular and supramarginal gyrus. The separate analysis of 1-back (alphasim cluster threshold = 82 voxels) revealed training-related activation changes in a cluster in right inferior parietal lobule including supramarginal gyrus and angular gryus.

When testing the 0-back and 3-back contrasts separately at p < .05 alphasim cluster corrected (alphasim cluster thresholds = 87 voxels), no significant group by time interaction effects were found. No training-related activity increases were detected in any of the performed analyses.

------------------------------Please insert Figure 3 approximately here----------------------------

------------------------------Please insert Table 3 approximately here----------------------------

Neural correlates of pre-post (t1-t2) changes in the Sternberg task

Training-specific activity changes from t1 to t2 were considered separately for the different operations in the Sternberg task (i.e. WM maintenance vs. updating) during the task delay phase (see Table 3).

The fMRI data indicate that interaction effects of group (control vs. training group) and time (t1 vs. t2) were only present during updating but not during maintenance (see Figure 3 panels B and C and Table 3; alphasim cluster thresholds: 64 voxels for updating and 58 voxels for maintenance). Specifically, one right lateral prefrontal cluster located in the middle frontal gyrus (MFG) extending to the caudal superior frontal sulcus (cSFS, BAs 6, 8, and 9) showed such an interaction effect during updating at p < .05 alphasim cluster corrected (see Table 3). As shown in Figure 3, panel B, the same activity pattern as in the combined update-3 & update-5 contrast was present in the update-3 as well as in the update-5 condition separately, however, only at an uncorrected threshold of p < .005, k > 10.

Neural overlap between training-related changes in n-back and Sternberg updating

To test for neural correlates of pre-post changes in the Sternberg task that were overlapping with n-back training effects, we computed a mask image that only included voxels that showed significant effects ($p < .05$ alphasim cluster corrected) in both the n-back and Sternberg tasks. The overlap mask included two clusters (see Table 3), one in the right MFG ($x = 27$, $y = 11$, $z = 55$), and one more lateral cluster in right MFG/ cSFS ($x = 36$, $y = 17$, $z = 46$). As there were significant baseline differences between groups at t1 in the more medial cluster in the n-back and in the Sternberg task further analyses were restricted to the more lateral frontal cluster

Percent signal change (PSC) values for the peak of the right MFG/cSFS cluster are depicted in the bar graph in Figure 3, panel C, separately for the combined 1-& 2-back task and combined update-3 & update-5 contrast. The ROI analysis illustrates strong decreases in PSC from t1 to t2 during 1-& 2-back and during updating for the training group, while no changes or changes in the opposite direction were present in the control group.

Note that significant group differences in activity at t1 within this ROI (see Table 4) were restricted to the update-5 condition of the Sternberg task - only here, higher PSC was present in the training group compared to the control group at t1. For all other conditions there were no significant baseline differences between training group and control group at t1, although several medium-sized effects ($d > .50$) of group on PSC indicate a trend for such differences. The central finding of selective, PSC decreases over time in the right MFG/cSFS in the training group in both tasks (see also Supplementary results, 'RT subgroup analysis'), suggests common effects of n-back training across tasks within this region, thus pointing towards training-related neural changes in the transfer task.

------------------------------Please insert Table 4 approximately here------------------------------

Post-hoc power analyses of fMRI-results

To test for appropriate power (1-beta > .80) of our main fMRI findings, we performed post-hoc power analyses. Comparing the training-related activity changes in the overlap peak (x = 36, y = 17, z = 46) between training and control group (see Figure 3, panel C) revealed an effect size of d = 1.12 in the n-back training task and d = 1.36 in the Sternberg transfer task. At an alpha error probability level of p = .05, the achieved power was .83 for n-back and .94 for Sternberg.

Correlations between pre-post changes in fMRI percent signal change (PSC) and behavior

In the n-back task, performance gains in the 2-back condition were correlated with pre-post activation changes (PSC) in the overlap peak (x = 36, y = 17, z = 46) in the 2-back contrast (r = -.44, p = .017, see Figure 4.) and in the combined 1- & 2-back contrast (r = -.40, p = .030) across the entire sample. This indicates that the magnitude of activation decrease was associated with performance increase in 2-back. However, no correlations between activation changes and n-back performance gains within each group were found (p's > .13).

To test if activation changes in the overlap peak in 1-back and in 2-back separately, as well as in the combined 1-& 2-back contrast are related to significant pre-post changes in behavioral transfer tasks we performed correlation analyses in tasks that showed significant improvements in the training but not in the control group (D2 Test, Digit Symbol Substitution, Stroop Interference, and LPS Figural Relations Test). In the entire sample, we found correlations between pre-post decrease in 2-back activation and pre-post increase in D2 Test (r = -.37, p = .049) and LPS Figural Relations Test (r = -.48, p = .009). Similarly, pre-post activation decreases in the combined 1-& 2-back contrast also correlated with D2 Test (r = -.37, p = .049) and LPS Figural Relations Test (r = -.40, p = .031). Within the training

25

group, a significant correlation was found between 1-&2-back activation decrease and Digit Symbol Substitution increase after training (r = -.53, p = .041). These results suggest that pre-post decreases in lateral MFG/cSFS may be associated with improvements in processing speed and fluid intelligence.

In the Sternberg task, pre-post activation changes in the combined update-3 & update-5 contrast did not reveal any significant correlations with performance gains (% correct) in the whole group (r = .12, p = .52), nor within each group (both p's > .13).

------------------------------Please insert Figure 4 approximately here------------------------------

Discussion

In the present study, we investigated training-related gains in two verbal WM tasks and associated changes in BOLD response in older participants. Our results indicate that after four weeks of WM training, participants in the training group improve in their performance in the trained n-back task (see also previously published results of a subsample of the current study (Heinzel et al., 2014a)). This was accompanied by training-related decreases in BOLD signal in lateral prefrontal cortex including MFG and cSFS (BA 6/8) as well as in medial frontal gyrus/mid-cingulate and supplementary motor area in the 1- & 2-back condition of the n-back task. Activity decreases in medial PFC and right lateral MFG/cSFS are congruent with previous findings of training-related activity decreases. In the untrained Sternberg task, we found a decrease of activity in the training group compared to the control group that was overlapping with the training-related activity decrease in the n-back task in right MFG/cSFS exclusively in the delay phase of the updating condition. Similar to the results in n-back, albeit to a smaller degree, performance in the Sternberg task and other behavioral transfer tasks increased after training.

26

With respect to the five specific aims outlined in the Introduction, these results support (i) the notion of a training-related increase in processing efficiency as postulated by the CRUNCH model. (ii) Training-related activation decreases that were found in 1-back and 2-back (low and medium task demand) and not in 3-back (high task demand) are in line with predictions from task demand to activation functions postulating a shift in task demand to activation functions with increasing WM load. Findings specific to the updating condition suggest that (iii) training-related changes in updating processes are involved in transfer effects and (iv) that these transfer effects are related to common activation decreases in a region in right lateral PFC. The question whether (v) this neuronal overlap is related to near and far behavioral transfer effects remains partly open since most effects were not specific to the training group. To our knowledge, this is the first study that shows neural correlates of a transfer effect within the working memory domain in older adults.

Behavioral training and transfer effects

The behavioral training results are in line with other behavioral WM training studies involving n-back in older adults (Heinzel et al., 2014b, 2014c; Li et al., 2008; Schmiedek et al., 2010) that reported similar training gains. It has to be noted that in addition to the group by time interaction, a main effect of time across both groups in the n-back task performance was found. This indicates that practicing the n-back task three times outside and twice inside the MRI in the control group may already have induced familiarization effects and led to small performance increases also in the control group.

In addition to behavioral gains, transfer effects to several untrained behavioral tasks outside the scanner were detected. The finding of performance increases in the untrained Stroop and Sternberg updating tasks support our hypothesis that n-back WM training may lead to an improvement in executive control processing and is in line with previous findings

27

in a subsample of this study (Heinzel et al., 2014a). Consistent with the model on executive functions by Miyake et al. (2000), it seems that transfer effects of n-back training generalize across executive functions and include both updating as well as inhibition components of executive functions.

No behavioral transfer effects were found in Digit Span, Digit Symbol, Verbal Fluency, and Raven's SPM. These results contribute to the ongoing debate on which transfer findings can be expected from WM training studies (Melby-Lervåg and Hulme, 2013; Redick et al., 2012; Shipstead et al., 2012; Slagter, 2012). Since the training procedure applied in this study improved performance in the n-back task, an increase in Digit Span performance (near transfer task) might have been expected. Contrary to this hypothesis, we did not find any transfer effect in this test. Digit Span Forward task performance is thought to rely on the amount of items a participant is able to store for the period of several seconds (Bopp and Verhaeghen, 2005). Recent meta-analyses have indicated medium-sized transfer effects to tasks within the WM domain ("near" transfer, (Melby-Lervåg and Hulme, 2013; Schwaighofer et al., 2015; Spencer-Smith and Klingberg, 2015). Still, an increase in performance in WM tasks does not necessarily imply an increase in digit span. As discussed by (Shipstead et al., 2012), despite a large number of studies, evidence for a training-related increase in short-term memory storage is still not fully convincing as investigating transfer to Digit Span Forward has produced mixed results (see also Schwaighofer et al., 2015). It may well be that many transfer effects to other WM tasks are mainly due to an increase in processing speed, and/or the development of more efficient strategies to update and inhibit certain information in WM. Also, no transfer to Digit Span Backward was detected, while performance in other executive tasks (Sternberg updating, Stroop Interference) increased. It seems that our n-back training may have improved specific executive subcomponents (updating and inhibition) rather than the manipulation subcomponent of WM or general short-

term memory storage in the current study. It has been argued that the manipulation subcomponent of WM which is required for Digit Span Backward, relies more on processes related to posterior parietal cortex (Champod and Petrides, 2007) as compared to more prefrontally-based executive processes. This view is further supported by training-related changes in lateral and medial frontal areas in n-back that overlap specifically with changes in the updating condition of the Sternberg task in the current study.

Conflicting results in the investigation of transfer to measures of fluid intelligence (transfer to LPS, but no transfer to Raven's SPM) might be attributed to different testing procedures applied. While item difficulty in LPS constantly increased when the test was carried out, the item difficulty in Raven's SPM was alternated between easy and difficult items. Therefore, an increase in processing speed performance as indicated by transfer to D2 test might have facilitated the completion of LPS Figural Relations, but not Raven's SPM. In fact, pre-post increase in D2 test scores was significantly correlated with pre-post performance changes in LPS Figural Relations (r = .50, p = .006), but not in Raven's SPM (r = .07, p = .704). However, this interpretation can only be speculative at this point and further investigations on mechanisms involved in transfer to different fluid intelligence measures are required.

Neural correlates of training and transfer effects

FMRI results at t1 showed mainly bilateral fronto-parietal activations in all n-back conditions as well as in the Sternberg task (see Figure 2, and Supplementary Tables S2 and S3), replicating previous findings of WM studies (for reviews, see (Nee et al., 2013; Owen et al., 2005). The findings of strong bilateral recruitment of fronto-parietal areas may indicate age-related effects as discussed in the "hemispheric asymmetry reduction in old age" (HAROLD) model by (Cabeza, 2002) and the CRUNCH model (Reuter-Lorenz and Cappell,

29

2008). While a hemispheric asymmetry towards the left hemisphere was found in younger adults in verbal WM tasks, older adults show a less asymmetric activation pattern, indicating compensational mechanisms (Reuter-Lorenz et al., 2000). When WM load was manipulated, younger adults showed a linear WM load-dependent increment in fronto-parietal activity while less WM load-dependent variation in fronto-parietal activity has been reported in older adults (Heinzel et al., 2014a; Nagel et al., 2011), also interpreted in the framework of age-related compensational activations and capacity limitations.

The t1 data were further used to check if unbalanced allocation to the different scanner across the two groups had an impact on the results. Note that scanner site was included as a covariate in all fMRI analyses to control for possible effects of the site. To further assess such potential effects, we additionally compared activity between training and control group at t1. No significant differences were found between groups in any fronto-parietal regions, when correcting for multiple comparisons. The only activation differences were detected in primary visual cortex in the maintenance conditions and the update-3-condition of the Sternberg task. This might be due to hardware differences in the stimulus presentation setups between the two sites. As these effects were cancelled out in pre-post analyses and data quality checks revealed no differences in mean PSC between sites ($PSC_{site1} = .37\%$; $PSC_{site2} = .38\%$), we are convinced that the pre-post results described in this study can be attributed to effects of the intervention rather than differences between sites. In addition, we excluded that differences in reaction times at t2 might underlie the training and transfer effects in the overlap ROI by showing that activity differences between training and control group were also present when reaction times were approximately matched between groups (results reported in the Supplement).

Taken together, the pre-post fMRI results of our study endorse the concept of a training-related increase in efficiency of the WM system as discussed by Lustig et al., (2009).

In the group of older training participants, activity during n-back in medial frontal gyrus/mid-cingulate (BA 6/32) and in right MFG/cSFS (BA 6/8) decreased from t1 to t2 selectively in the training group and not in the control group. Training-related reduction in medial frontal gyrus/mid-cingulate may be associated with reduced attentional effort for monitoring response conflict and errors (Ridderinkhof et al., 2004) after the training program. Training-related activity decreases in right MFG/cSFS in both n-back and Sternberg updating tasks may reflect that compensatory recruitment of additional brain areas in the right hemisphere, as postulated by the HAROLD model (Cabeza, 2002) partially recedes through the training procedure (Erickson et al., 2007; Lustig et al., 2009).

Our results of common activity decreases in low and medium (1- and 2-back) WM load conditions in the n-back task and Sternberg updating task support predictions from the CRUNCH model (Reuter-Lorenz and Cappell, 2008), suggesting that less neural resources are required to achieve the same or even better performance in the n-back task as well as the Sternberg updating condition after training. According to the CRUNCH model, less or no training-related decreases in BOLD response would be expected for very difficult tasks. It seems that in the Dahlin et al., (2008) study, a reason for the absence of neural overlap between an updating training task and a 3-back transfer task in the older group may have been the high difficulty of both training and transfer tasks. While older and younger participants in the Dahlin et al., (2008) study were able to increase their updating performance through training of a letter memory task, transfer effects to a 3-back task were only present in the younger adults. Dahlin and colleagues discussed their finding of an increased activity of the striatum with training as a neural correlate of the regulation of "updating in WM by affecting dopaminergic modulation of the prefrontal cortex" (Dahlin et al., 2008, p.1512). This seems plausible in the context of the conceptions of Frank and colleagues (Frank et al., 2001) and also the findings described by (McNab and Klingberg, 2008) according to which the striatum

31

is involved in gating information to the prefrontal cortex during encoding (but see (Kühn et al., 2013). Dahlin and colleagues (Dahlin et al., 2008) related the missing transfer effects in the older participants in their study to a deficient functioning of the striatum in old age. The present study shows that neural transfer effects can be observed in older participants, but that they are related primarily to activity changes in the right MFG/cSFS. Note that the applied delayed recognition task was special in its sensitivity to different WM component processes. We showed that the main overlap between the n-back training effects and the Sternberg effects was in the right MFG/cSFS during WM updating. This region is part of the dorsal attention system (for review, see Nee et al., 2013) and has been associated with attentional control processes involved in WM in several studies (Curtis and D'Esposito, 2003; Gazzaley and Nobre, 2012; Kok et al., 2006). Therefore, the activity changes observed in the present study may have been related to decreasing attentional control processes required to cope with the tasks after training, rather than to training the gating signal from the striatum. Accordingly, our results support the concept of a training-related increase in processing efficiency, i.e. requiring less neural resources while gaining higher behavioral performance. Of note, this effect goes along with a decrease in the need for compensatory mechanisms as predicted by the HAROLD model, reflected by reduced lateralization, suggesting a plausible trade-off between increased efficiency and compensatory processes. One hypothesis put forth from our findings is that indeed such trade-offs generalize across executive functions, as reflected by the neuronal and behavioral transfer effects found.

Brain-behavior correlations

Results from brain-behavior correlations indicate that pre-post activation decreases in the 2-back and in the combined 1- & 2-back contrast were related to pre-post performance increases in 2-back, D2 Test, and LPS Figural Relations Test. Therefore, increased processing

efficiency, as proposed by activation decreases, might manifest in higher accuracy at medium difficulty level of n-back as well as in processing speed and fluid intelligence tasks. However, most brain-behavior relationships were only found in the entire sample, and vanished when investigating each group separately. Therefore, these associations seem to be not specific to the effects of training. Within the training group, only the correlation with Digit Symbol Substitution was found significant, suggesting that training-related brain activation changes in working memory may be associated with a performance increase in this attention-related processing speed task. No correlation between behavior in the Sternberg transfer task and the activity in the overlap ROI in right lateral MFG/cSFS were present. Note that missing links of training-related behavioral changes and changes in brain activity, does not necessarily indicate that there was no functional relevance of the changes in brain activity. Changes in brain activity might affect behavior in numerous ways. For example, they might be part of changes in functional connectivity patterns, which are known to be crucial for learning-related plasticity (Greenwood, 2007), but weren't assessed in the current study. Also, they might reflect the sensitivity of imaging techniques to detect changes that are actually not yet expressed in behavior. Accordingly, in our view those activity changes can be considered as relevant even in the absence of correlations with behavioral changes.

Limitations and future research

Some limitations need to be considered when interpreting the results of the current study. Firstly, the sample sizes were relatively small, group allocation was not random, and education level was comparatively high, thus generalizing results to the general population requires replication in larger, randomized, and more representative samples. Larger samples may allow investigating inter-individual differences in training gains and associated neural changes in future studies to further understand which participants benefit from this type of

33

cognitive training and who does not (see Brehmer et al., 2014 for a discussion on the topic).

Further studies with larger sample sizes would also allow assessing the effects of baseline

differences in more detail. Here, we obtained some trends (i.e. medium-sized effects) for

differences in activity at t1 between training and control group, which, however, did not reach

significance. It can be speculated that the randomization of training and control groups instead

of the matching procedure applied in the current study would have prevented baseline

differences. Larger samples would allow building subgroups that are more precisely matched

with regard to baseline activity in order to test the actual effects of such differences on

training and transfer effects. Also, behavioral transfer effects to the Sternberg task were

mostly present at trend levels only, although effect size analyses and post-hoc t-tests provided

evidence for significant medium-sized behavioral effects in the training group but not in the

control group. Note, that on the behavioral level some trends for training-specific changes

were also present in the maintenance condition. Still, the obtained training-related changes in

the brain data for the updating condition, but not for the maintenance condition indicate that

transfer was potentially related specifically to the updating component of the n-back training.

Further research is required to disentangle the contributions of both operations in more detail.

Secondly, we only included a no-contact control group, therefore social interaction or other

unspecific effects associated with the training procedure might have influenced both

behavioral and neural changes in the training group. In future studies, active control groups

should be included in the study design. Nonetheless, as no general training-related benefit for

all transfer tasks was observed, training-effects do not seem to be solely related to unspecific

effects. Recent research investigating effects of different types of interventions on cognitive

performance in aging indicates that not only cognitive training but also physical exercise may

have beneficial effects (Angevaren et al., 2008; Voelcker-Rehage and Niemann, 2013). An

interesting focus of future research could be to compare different physical and cognitive

34

training regimes as well as combinations of both types, as suggested by a recent review on cognitive and physical interventions (Bamidis et al., 2014).

Conclusion

After 4 weeks of n-back training, BOLD signal in the right MFG/cSFS decreased in the training group in both the trained n-back and the untrained Sternberg updating task. Behavioral transfer to processing speed, executive functions, and figural relations (fluid intelligence) was found. Performance gains in these untrained tasks suggest that transfer to other cognitive domains may remain possible throughout the lifespan. FMRI findings indicate a training-related increase in processing efficiency of WM networks, specifically related to the process of WM updating.

Disclosure statement

All authors and their institutions declare to have no actual or potential conflicts of interest to disclose.

References

Ackerman, P.L., Beier, M.E., Boyle, M.O., 2005. Working memory and intelligence: the same or different constructs? Psychol. Bull. 131, 30–60. doi:10.1037/0033-2909.131.1.30

Angevaren, M., Aufdemkampe, G., Verhaar, H.J.J., Aleman, A., Vanhees, L., 2008. Physical activity and enhanced fitness to improve cognitive function in older people without known cognitive impairment. Cochrane Database Syst. Rev. CD005381. doi:10.1002/14651858.CD005381.pub2

Baddeley, A., 2003. Working memory: looking back and looking forward. Nat. Rev. Neurosci. 4, 829–839. doi:10.1038/nrn1201

Baddeley, A., 2000. The episodic buffer: a new component of working memory? Trends Cogn. Sci. 4, 417–423. doi:10.1016/S1364-6613(00)01538-2

Bamidis, P.D., Vivas, A.B., Styliadis, C., Frantzidis, C., Klados, M., Schlee, W., Siountas, A., Papageorgiou, S.G., 2014. A review of physical and cognitive interventions in aging. Neurosci. Biobehav. Rev., Applied Neuroscience: Models, methods, theories, reviews. A Society of Applied Neuroscience (SAN) special issue. 44, 206–220. doi:10.1016/j.neubiorev.2014.03.019

Barulli, D., Stern, Y., 2013. Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. Trends Cogn. Sci. 17, 502–509. doi:10.1016/j.tics.2013.08.012

Bopp, K.L., Verhaeghen, P., 2005. Aging and Verbal Memory Span: A Meta-Analysis. J. Gerontol. B. Psychol. Sci. Soc. Sci. 60, P223–P233. doi:10.1093/geronb/60.5.P223

Brehmer, Y., Kalpouzos, G., Wenger, E., Lövdén, M., 2014. Plasticity of brain and cognition in older adults. Psychol. Res. 78, 790–802. doi:10.1007/s00426-014-0587-z

Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., Bäckman, L., 2011. Neural correlates of training-related working-memory gains in old age. NeuroImage 58, 1110–1120. doi:10.1016/j.neuroimage.2011.06.079

Burgaleta, M., Colom, R., 2008. Short-term storage and mental speed account for the relationship between working memory and fluid intelligence. Psicothema 20, 780–785.

Buschkuehl, M., Jaeggi, S.M., Jonides, J., 2012. Neuronal effects following working memory training. Dev. Cogn. Neurosci. 2 Suppl 1, S167–179. doi:10.1016/j.dcn.2011.10.001

Cabeza, R., 2002. Hemispheric asymmetry reduction in older adults: the HAROLD model. Psychol. Aging 17, 85–100.

Cabeza, R., Anderson, N.D., Locantore, J.K., McIntosh, A.R., 2002. Aging gracefully: compensatory brain activity in high-performing older adults. NeuroImage 17, 1394–1402.

Cabeza, R., Daselaar, S.M., Dolcos, F., Prince, S.E., Budde, M., Nyberg, L., 2004. Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval. Cereb. Cortex N. Y. N 1991 14, 364–375.

Champod, A.S., Petrides, M., 2007. Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. Proc. Natl. Acad. Sci. U. S. A. 104, 14837–14842. doi:10.1073/pnas.0607101104

Chein, J.M., Fiez, J.A., 2010. EVALUATING MODELS OF WORKING MEMORY THROUGH THE EFFECTS OF CONCURRENT IRRELEVANT INFORMATION. J. Exp. Psychol. Gen. 139, 117–137. doi:10.1037/a0018200

Chen, T., Li, D., 2007. The roles of working memory updating and processing speed in mediating age-related differences in fluid intelligence. Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn. 14, 631–646. doi:10.1080/13825580600987660

Clay, O.J., Edwards, J.D., Ross, L.A., Okonkwo, O., Wadley, V.G., Roth, D.L., Ball, K.K., 2009. Visual function and cognitive speed of processing mediate age-related decline in memory span and fluid intelligence. J. Aging Health 21, 547–566. doi:10.1177/0898264309333326

Cohen, J.D., Perlstein, W.M., Braver, T.S., Nystrom, L.E., Noll, D.C., Jonides, J., Smith, E.E., 1997. Temporal dynamics of brain activation during a working memory task. Nature 386, 604–608. doi:10.1038/386604a0

Collette, F., Van der Linden, M., 2002. Brain imaging of the central executive component of working memory. Neurosci. Biobehav. Rev. 26, 105–125.

Conway, A.R.A., Kane, M.J., Engle, R.W., 2003. Working memory capacity and its relation to general intelligence. Trends Cogn. Sci. 7, 547–552.

Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. Trends Cogn. Sci. 7, 415–423.

Dahlin, E., Neely, A.S., Larsson, A., Bäckman, L., Nyberg, L., 2008. Transfer of learning after updating training mediated by the striatum. Science 320, 1510–1512. doi:10.1126/science.1155466

D'Esposito, M., Detre, J.A., Alsop, D.C., Shin, R.K., Atlas, S., Grossman, M., 1995. The neural basis of the central executive system of working memory. Nature 378, 279–281. doi:10.1038/378279a0

D'Esposito, M., Postle, B.R., Rypma, B., 2000. Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. Exp. Brain Res. Exp. Hirnforsch. Expérimentation Cérébrale 133, 3–11.

Doumas, M., Rapp, M.A., Krampe, R.T., 2009. Working Memory and Postural Control: Adult Age Differences in Potential for Improvement, Task Priority, and Dual Tasking. J. Gerontol. B. Psychol. Sci. Soc. Sci. 64B, 193–201. doi:10.1093/geronb/gbp009

Erickson, K.I., Colcombe, S.J., Wadhwa, R., Bherer, L., Peterson, M.S., Scalf, P.E., Kim, J.S., Alvarado, M., Kramer, A.F., 2007. Training-induced plasticity in older adults: effects of training on hemispheric asymmetry. Neurobiol. Aging 28, 272–283. doi:10.1016/j.neurobiolaging.2005.12.012

Eyler, L.T., Sherzai, A., Kaup, A.R., Jeste, D.V., 2011. A Review of Functional Brain Imaging Correlates of Successful Cognitive Aging. Biol. Psychiatry 70, 115–122. doi:10.1016/j.biopsych.2010.12.032

Frank, M.J., Loughry, B., O'Reilly, R.C., 2001. Interactions between frontal cortex and basal ganglia in working memory: a computational model. Cogn. Affect. Behav. Neurosci. 1, 137–160.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189–210. doi:10.1002/hbm.460020402

Fuster, J.M., 2004. Upper processing stages of the perception-action cycle. Trends Cogn. Sci. 8, 143–145.

Gazzaley, A., Nobre, A.C., 2012. Top-down modulation: Bridging selective attention and working memory. Trends Cogn. Sci. 16, 129–135. doi:10.1016/j.tics.2011.11.014

Grady, C., 2012. The cognitive neuroscience of ageing. Nat. Rev. Neurosci. 13, 491–505. doi:10.1038/nrn3256

Gray, J.R., Chabris, C.F., Braver, T.S., 2003. Neural mechanisms of general fluid intelligence. Nat. Neurosci. 6, 316–322. doi:10.1038/nn1014

Greenwood, P.M., 2007. Functional plasticity in cognitive aging: review and hypothesis. Neuropsychology 21, 657–673. doi:10.1037/0894-4105.21.6.657

Heinzel, S., Lorenz, R.C., Brockhaus, W.-R., Wüstenberg, T., Kathmann, N., Heinz, A., Rapp, M.A., 2014a. Working memory load-dependent brain response predicts behavioral training gains in older adults. J. Neurosci. Off. J. Soc. Neurosci. 34, 1224–1233. doi:10.1523/JNEUROSCI.2463-13.2014

Heinzel, S., Riemer, T.G., Schulte, S., Onken, J., Heinz, A., Rapp, M.A., 2014b. Catechol-O-methyltransferase (COMT) Genotype Affects Age-Related Changes in Plasticity in Working Memory: A Pilot Study. BioMed Res. Int. 2014. doi:10.1155/2014/414351

Heinzel, S., Schulte, S., Onken, J., Duong, Q.-L., Riemer, T.G., Heinz, A., Kathmann, N., Rapp, M.A., 2014c. Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn. 21, 146–173. doi:10.1080/13825585.2013.790338

Jonides, J., 2004. How does practice makes perfect? Nat. Neurosci. 7, 10–11. doi:10.1038/nn0104-10

Klingberg, T., 2010. Training and plasticity of working memory. Trends Cogn. Sci. 14, 317–324. doi:10.1016/j.tics.2010.05.002

Kok, A., Ridderinkhof, K.R., Ullsperger, M., 2006. The control of attention and actions: Current research and future developments. Brain Res. 1105, 1–6. doi:10.1016/j.brainres.2006.03.027

Kühn, S., Schmiedek, F., Noack, H., Wenger, E., Bodammer, N.C., Lindenberger, U., Lövden, M., 2013. The dynamics of change in striatal activity following updating training. Hum. Brain Mapp. 34, 1530–1541. doi:10.1002/hbm.22007

Kyllonen, P.C., Christal, R.E., Oktober. Reasoning ability is (little more than) working-memory capacity?! Intelligence 14, 389–433. doi:10.1016/S0160-2896(05)80012-1

Linden, D.E.J., 2007. The working memory networks of the human brain. Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry 13, 257–267. doi:10.1177/1073858406298480

Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., Lindenberger, U., 2008. Working memory plasticity in old age: practice gain, transfer, and maintenance. Psychol. Aging 23, 731–742. doi:10.1037/a0014343

Lustig, C., Shah, P., Seidler, R., Reuter-Lorenz, P.A., 2009. Aging, training, and the brain: a review and future directions. Neuropsychol. Rev. 19, 504–522. doi:10.1007/s11065-009-9119-9

McNab, F., Klingberg, T., 2008. Prefrontal cortex and basal ganglia control access to working memory. Nat. Neurosci. 11, 103–107. doi:10.1038/nn2024

Melby-Lervåg, M., Hulme, C., 2013. Is working memory training effective? A meta-analytic review. Dev. Psychol. 49, 270–291. doi:10.1037/a0028228

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202. doi:10.1146/annurev.neuro.24.1.167

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D., 2000. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cognit. Psychol. 41, 49–100. doi:10.1006/cogp.1999.0734

Mohr, H.M., Goebel, R., Linden, D.E.J., 2006. Content- and task-specific dissociations of frontal activity during maintenance and manipulation in visual working memory. J. Neurosci. Off. J. Soc. Neurosci. 26, 4465–4471. doi:10.1523/JNEUROSCI.5232-05.2006

Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., Mellits, E.D., Clark, C., 1989. The Consortium to Establish a Registry for Alzheimer's Disease

(CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 39, 1159–1165.

Nagel, I.E., Preuschhof, C., Li, S.-C., Nyberg, L., Bäckman, L., Lindenberger, U., Heekeren, H.R., 2011. Load modulation of BOLD response and connectivity predicts working memory performance in younger and older adults. J. Cogn. Neurosci. 23, 2030–2045. doi:10.1162/jocn.2010.21560

Nee, D.E., Brown, J.W., Askren, M.K., Berman, M.G., Demiralp, E., Krawitz, A., Jonides, J., 2013. A meta-analysis of executive components of working memory. Cereb. Cortex N. Y. N 1991 23, 264–282. doi:10.1093/cercor/bhs007

Noack, H., Lövdén, M., Schmiedek, F., Lindenberger, U., 2009. Cognitive plasticity in adulthood and old age: gauging the generality of cognitive intervention effects. Restor. Neurol. Neurosci. 27, 435–453. doi:10.3233/RNN-2009-0496

Nyberg, L., Dahlin, E., Stigsdotter Neely, A., Bäckman, L., 2009. Neural correlates of variable working memory load across adult age and skill: dissociative patterns within the fronto-parietal network. Scand. J. Psychol. 50, 41–46. doi:10.1111/j.1467-9450.2008.00678.x

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., Bäckman, L., 2012. Memory aging and brain maintenance. Trends Cogn. Sci. 16, 292–305. doi:10.1016/j.tics.2012.04.005

Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25, 46–59. doi:10.1002/hbm.20131

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56, 303–308.

Rajah, M.N., D'Esposito, M., 2005. Region-specific changes in prefrontal function with age: a review of PET and fMRI studies on working and episodic memory. Brain J. Neurol. 128, 1964–1983. doi:10.1093/brain/awh608

Redick, T.S., Shipstead, Z., Harrison, T.L., Hicks, K.L., Fried, D.E., Hambrick, D.Z., Kane, M.J., Engle, R.W., 2012. No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study. J. Exp. Psychol. Gen. doi:10.1037/a0029082

Reuter-Lorenz, P.A., Cappell, K.A., 2008. Neurocognitive Aging and the Compensation Hypothesis. Curr. Dir. Psychol. Sci. 17, 177–182. doi:10.1111/j.1467-8721.2008.00570.x

Reuter-Lorenz, P.A., Jonides, J., Smith, E.E., Hartley, A., Miller, A., Marshuetz, C., Koeppe, R.A., 2000. Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. J. Cogn. Neurosci. 12, 174–187.

Richmond, L.L., Morrison, A.B., Chein, J.M., Olson, I.R., 2011. Working memory training and transfer in older adults. Psychol. Aging 26, 813–822. doi:10.1037/a0023631

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., Nieuwenhuis, S., 2004. The Role of the Medial Frontal Cortex in Cognitive Control. Science 306, 443–447. doi:10.1126/science.1100301

Salthouse, T.A., 1996. The processing-speed theory of adult age differences in cognition. Psychol. Rev. 103, 403–428.

Schmiedek, F., Lövdén, M., Lindenberger, U., 2010. Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. Front. Aging Neurosci. 2. doi:10.3389/fnagi.2010.00027

Schneider-Garces, N.J., Gordon, B.A., Brumback-Peltz, C.R., Shin, E., Lee, Y., Sutton, B.P., Maclin, E.L., Gratton, G., Fabiani, M., 2010. Span, CRUNCH, and beyond: working

memory capacity and the aging brain. J. Cogn. Neurosci. 22, 655–669. doi:10.1162/jocn.2009.21230

Schwaighofer, M., Fischer, F., Bühner, M., 2015. Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators. Educ. Psychol. 50, 138–166. doi:10.1080/00461520.2015.1036274

Shipstead, Z., Redick, T.S., Engle, R.W., 2012. Is working memory training effective? Psychol. Bull. 138, 628–654. doi:10.1037/a0027473

Slagter, H.A., 2012. Conventional working memory training may not improve intelligence. Trends Cogn. Sci. doi:10.1016/j.tics.2012.10.001

Spencer-Smith, M., Klingberg, T., 2015. Benefits of a Working Memory Training Program for Inattention in Daily Life: A Systematic Review and Meta-Analysis. PLoS ONE 10. doi:10.1371/journal.pone.0119522

Veltman, D.J., Rombouts, S.A.R.B., Dolan, R.J., 2003. Maintenance versus manipulation in verbal working memory revisited: an fMRI study. NeuroImage 18, 247–256.

Voelcker-Rehage, C., Niemann, C., 2013. Structural and functional brain changes related to different types of physical activity across the life span. Neurosci. Biobehav. Rev. 37, 2268–2295. doi:10.1016/j.neubiorev.2013.01.028

Zelinski, E.M., 2009. Far transfer in cognitive training of older adults. Restor. Neurol. Neurosci. 27, 455–471. doi:10.3233/RNN-2009-0495

Table legends

Table 1:   Means, standard errors of the mean, effect sizes (Cohen's d), and group by time analyses of variance (F-values, df, and p-values) of the performance (hitrate minus false alarm rate) in the n-back task (**panel A**) and the performance (% correct) in the Sternberg task (**panel B**) in the training and control group at pre (t1) and post assessment (t2).

Table 2:   Means, standard errors of the mean, and effect sizes (Cohen's d) of the performance (% correct) for all neuropsychological tests in the training and control group at pre (t1) and post assessment (t2), as well as p-values and effect sizes (partial η^2) of the 2(group) by 2(time) analyses of variance for each test.

Table 3:   Regions, MNI coordinates, t-values, and cluster sizes of significant whole-brain results of training-related activity changes (group [training group > control group] by time [t1 > t2] interaction) in n-back and Sternberg tasks. Only clusters above a cluster size that yielded an alphasim correction threshold of $p < .05$, corrected, are reported. Hem = Hemisphere; BA = Brodmann area.

Table 4:   Means, standard errors of the mean, and effect sizes (Cohen's d) of percent signal change (PSC) derived from MarsBaR toolbox in the overlap peak (x = 36, y = 17, z = 46) in the right middle frontal gyrus/ caudal superior frontal sulcus [MFG/cSFS] for each condition of the n-back (**panel A**) and Sternberg tasks (**panel B**) in the training and control group at pre (t1) and post assessment (t2).

Figure legends

Figure 1: **Panel A:** N-back paradigm, example: 2-back; % correct (hitrate minus false alarm rate) in all n-back conditions at pre- (t1) and post-test (t2) in training and control group. **Panel B:** Sternberg paradigm, upper stream: maintenance task condition, lower stream: updating task condition; % correct for all Sternberg conditions at t1 and t2 in trainng and control group.

Figure 2: Main effects of task phase at pre-test (t1) in both groups at p < .05 (Family-wise error [FWE-] corrected for whole brain), k > 10 voxels; **upper panel**: 0-back, 1-back, 2-back, 3-back, and combined 1- & 2-back vs. implicit baseline; **middle panel**: Sternberg maintenance-3, maintenance-5, updating-3, updating-5, and combined updating-3 & -5 vs. implicit baseline; **lower panel**: Sternberg updating vs. maintenance; load 5 vs. load 3 contrasts.

Figure 3: **Panel A:** Whole-brain results of training-related activity changes in n-back (group [training group > control group] by time [t1 > t2]) for 1-back and 2-back separately, and the combined 1- & 2-back vs. implicit baseline contrast displayed at p < .005, k > 10 voxels; **panel B**: Whole-brain results of activation changes during the task delay phase of the Sternberg updating task (group [training group > control group] by time [t1 > t2]) for updating-3 and updating-5 separately, and the combined updating-3 & -5 vs. implicit baseline contrast, displayed at p < .005, k > 10 voxels; **panel C, left section:** n-back results are displayed in yellow, Sternberg updating results in red, and the overlapping regions are displayed in orange; **panel C, right section:** Percent signal change (PSC) in the overlap peak voxel (x = 36, y = 17, z = 46) in the right middle frontal gyrus/ caudal superior frontal sulcus [MFG/cSFS] in training and

control group at t1 and t2 for the combined 1-& 2-back contrast and for the combined Sternberg updating-3 & -5 contrast. Results show that activity decreased in the training group from t1 to t2, but did not change in the control group.

Figure 4: Correlation of performance gain in 2-back with pre-post changes (t2 > t1) in percent signal change (PSC) in the overlap peak voxel (x = 36, y = 17, z = 46) in 2-back.

A: Training Task: n-back Task
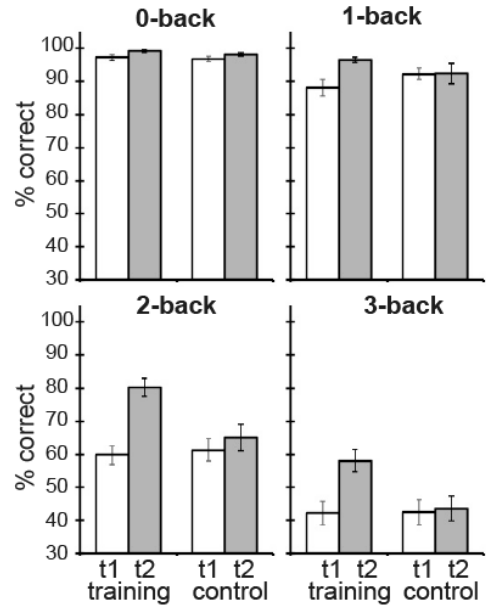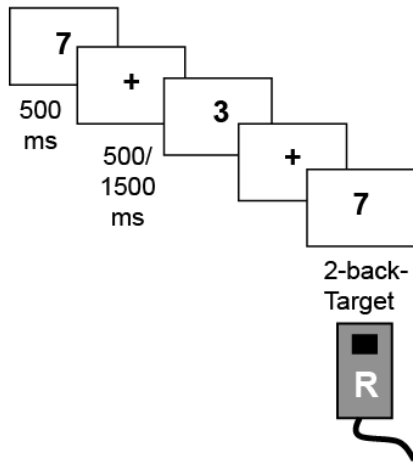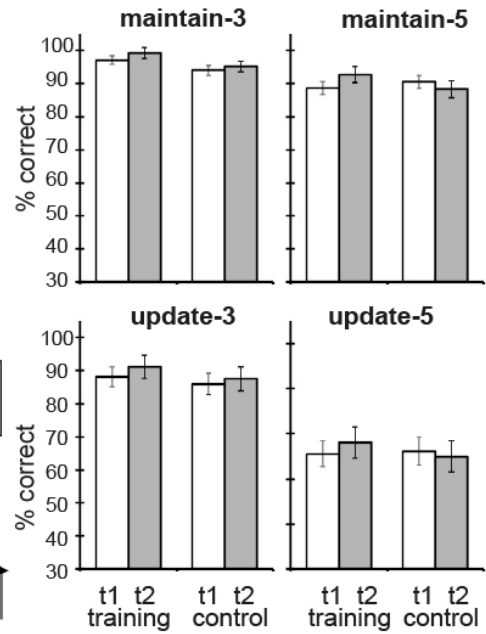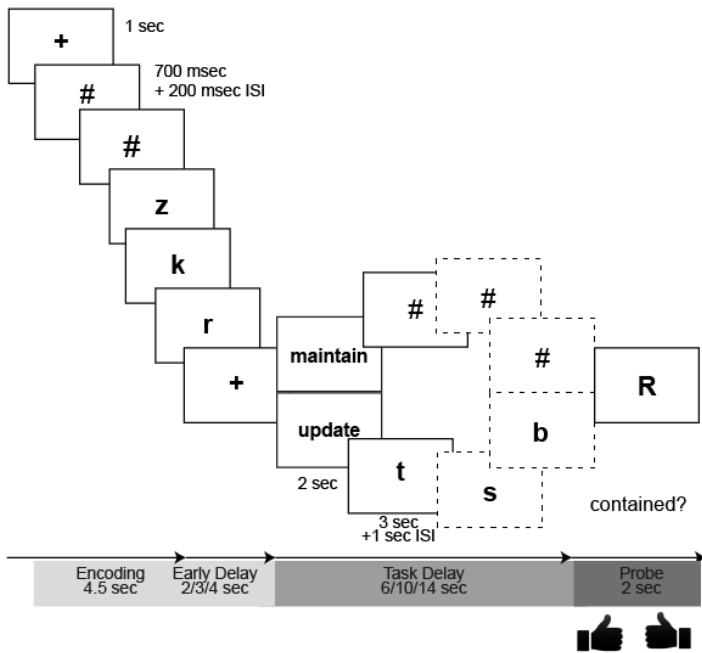

B: Transfer Task: Delayed Recognition ("Sternberg")

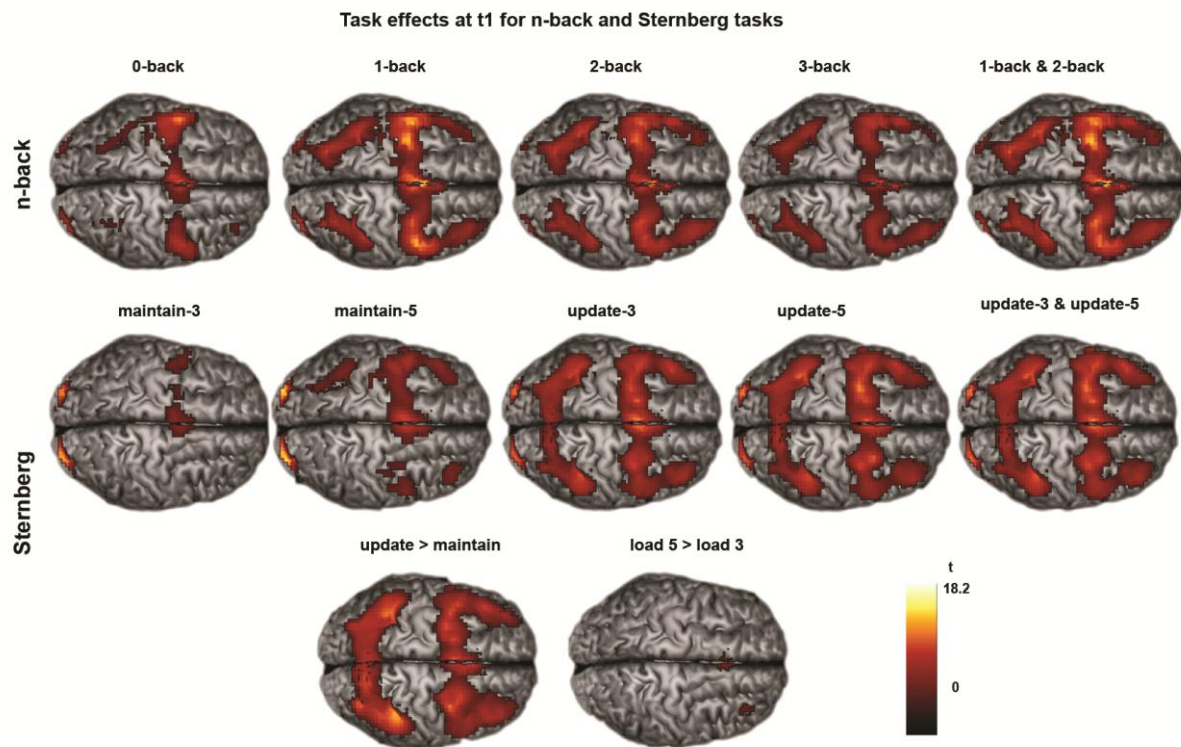**Figure 1**

Task effects at t1 for n-back and Sternberg tasks

Figure 2

Figure 3

Figure 4

Table 1

| | T1 | | | T2 | | | T2 > T1 | | Interaction Group |
	Training	Control	d	Training	Control	d	$d_{training}$	$d_{control}$	x Time
Panel A: n-back performance									
0-back	97.3 (0.9)	96.8 (0.9)	.15	99.2 (0.4)	98.2 (0.5)	.59	.73	.49	$F(1,27) = .13$, p = .717
1-back	88.1 (2.5)	92.3 (1.8)	-.50	96.5 (0.8)	92.4 (3.1)	.48	.97**	.01	$F(1,27) = 5.87$, p = .022
2-back	59.8 (2.9)	61.3 (3.4)	-.13	80.3 (2.8)	65.1 (4.0)	1.16**	1.86**	.26	$F(1,27) = 22.13$, p < .001
3-back	42.2 (3.5)	42.6 (3.8)	-.03	58.1 (3.5)	43.7 (3.8)	1.04**	1.19**	.08	$F(1,27) = 9.84$ p = .004
1- & 2-back	73.9 (2.0)	76.8 (2.2)	-.36	88.4 (1.6)	78.8 (3.4)	.97*	1.99**	.15	$F(1,27) = 20.50$, p < .001
Panel B: Sternberg task % correct									
maintain-3	97.2 (1.4)	94.0 (1.5)	.56	99.2 (1.6)	95.2 (1.6)	.65	.81	.15	$F(1,27) = 0.05$, p = .82
maintain-5	88.6 (1.9)	90.5 (2.0)	-.25	92.8 (2.5)	88.4 (2.6)	.46	.50*	-.22	$F(1,27) = 4.92$, p = .035
update-3	88.1 (3.0)	86.0 (3.2)	.17	91.2 (3.5)	87.5 (3.6)	.33	.42	.09	$F(1,27) = 0.46$, p = .51
update-5	75.6 (2.9)	76.2 (3.0)	-.06	80.0 (3.4)	75.0 (3.5)	.38	.35	-.10	$F(1,27) = 1.88$, p = .18
update-3 & update-5	81.8 (2.7)	81.1 (2.8)	.07	86.0 (2.6)	81.3 (3.8)	.39	.41*	.01	$F(1,27) = 1.40$, p = .23

* p < .05 ** p < .01

48

Table 2

	T1			T2			T2 > T1		group x time
	Training	Control	d	Training	Control	d	$d_{training}$	$d_{control}$	p (partial η²)
Digit Span Forward	6.27	6.86	-.55	6.60	6.71	-.09	.32	-.11	.340 (.03)
	(.27)	(.29)		(.27)	(.37)				
Digit Span Backward	4.80	4.71	.08	5.20	5.14	.05	.34	.40	.937 (.00)
	(.30)	(.29)		(.31)	(.29)				
D2 Test	403.47	421.64	-.22	443.60	419.79	.28	**.48****	-.02	**.003 (.28)****
	(20.95)	(23.29)		(21.71)	(22.46)				
Digit Symbol	31.47	32.57	-.17	33.87	34.57	-.09	**.32****	.26	.730 (.00)
Substitution	(1.58)	(1.89)		(1.87)	(2.12)				
Verbal Fluency	15.33	15.43	-.02	16.87	15.71	.23	.31	.06	.413 (.03)
	(1.29)	(1.10)		(1.24)	(1.38)				
Stroop Interference[b]	71.54	63.56	.47	63.66	62.97	.05	**-.41****	-.04	**.019 (.19)***
	(5.08)	(3.75)		(4.19)	(3.10)				
Raven's Standard	17.40	15.79	-.35	17.73	16.43	.31	.07	.14	.807 (.00)
Progr. Matrices	(1.17)	(1.28)		(1.20)	(1.00)				
LPS Figural	19.13	20.93	-.37	22.67	21.86	.18	**.77****	.18	**.043 (.14)***
Relations Test	(1.25)	(1.28)		(.87)	(1.44)				

Note:

[a] if not otherwise noted, mean values indicate number of correct items   * $p < .05$ ** $p < .01$

[b] Stroop interference = time [in sec] for interference task – [(color task * word task) / (color task + word task)]

Table 3

Contrast / Region	Hem	BA	MNI Coordinates x	y	z	t-value	cluster size
0-back (k > 86, alphasim-corr.)			No significant clusters.				
1-back (k > 81, alphasim-corr.)							
Inferior Parietal Lobule/ Supramarginal	R	40	44	-53	36	3.45	113
Gyrus/ Angular Gyrus			54	-46	36	3.42	included
			44	-39	42	3.21	included
2-back (k > 85, alphasim-corr.)							
Medial Frontal Gyrus/ Anterior Cingulate	R	32/ 9/	14	-26	46	4.03	251
Gyrus/ Mid-Cingulate Gyrus/		31/ 6/	11	20	42	3.80	included
Supplementary Motor Area/ Paracentral		8/ 24	14	17	36	3.69	included
Lobule/ Superior Frontal Gyrus							
Mid-Cingulate Gyrus/ Paracentral Lobule/	L	24/ 31/	-9	-26	49	3.77	99
Medial Frontal Gyrus/ Medial Frontal			-12	-13	46	3.77	included
Gyrus/ Supplementary Motor Area			-12	0	42	3.28	included
Superior Temporal Gyrus/ Middle	L	39/ 40	-52	-56	23	3.53	86
Temporal Gyrus/ Angular Gyrus/			-38	-56	16	3.51	included
Supramarginal Gyrus/			-48	-49	32	3.01	included
3-back (k > 86, alphasim-corr.)			No significant clusters.				
1-& 2-back (k >90, alphasim-corr.)							
Medial Frontal Gyrus/ Anterior Cingulate	R/L	32/ 6/	11	20	42	4.07	166
Gyrus/ Mid-Cingulate Gyrus/		8/ 9	8	30	36	3.75	included
Supplementary Motor Area			-9	23	46	3.46	included
Middle Frontal Gyrus/ Superior Frontal	R	6/ 8/ 9	24	7	56	3.92	140
Gyrus/			18	17	56	3.89	included
			34	20	39	3.18	included
Supramarginal Gyrus/ Interior Parietal	R	40	47	-53	32	3.43	112
Lobule/ Angular Gyurs			57	-46	39	3.21	included
			37	-49	36	3.17	included
Sternberg maintenance 3 & 5 (k > 57, alphasim-corr.)			No significant clusters.				
Sternberg updating 3 & 5 (k > 63, alphasim-corr.)							
Middle Frontal Gyrus/ Superior Frontal	R	8/ 6/ 9	36	17	46	4.16	68
Gyrus			27	11	55	3.42	included
			21	14	49	3.16	included
Sternberg updating 3 & 5 - overlap with 1-&2-back							
Middle Frontal Gyrus	R	9/ 8	36	17	46	4.16	22
Superior Frontal Gyrus/ Middle Frontal	R	6/ 8	27	11	55	3.42	23
Gyrus			21	14	49	3.16	included

50

Table 4

	T1			T2			T2 > T1	
	Training	Control	d	Training	Control	d	$d_{training}$	$d_{control}$
Panel A: n-back								
0-back	.089 (.073)	-.065 (.028)	.716	-.177 (.110)	-.088 (.022)	- .286	.726*	.248
1-back	.182 (.067)	.023 (.060)	.651	-.101 (.103)	-.018 (.044)	- .269	.813**	.197
2-back	.283 (.087)	.080 (.085)	.622	.049 (.114)	.124 (.061)	- .209	.572*	-.149
3-back	.256 (.061)	.118 (.081)	.508	.051 (.165)	.123 (.071)	- .146	.410	-.018
1- & 2-back	.233 (.068)	.052 (.069)	.695	-.026 (.096)	.053 (.050)	- .265	.762**	-.006
Panel B: Sternberg								
maintain-3	-.037 (.038)	-.86 (.022)	.404	-.052 (.164)	-.058 (.038)	.051	.090	-.186
maintain-5	.008 (.030)	-.027 (.024)	.399	-.042 (.027)	-.006 (.023)	-.366	.450	-.235
update-3	.144 (.052)	.075 (.043)	.380	.037 (.041)	.121 (.043)	-.527	.566**	-.289
update-5	.200 (.054)	.056 (.038)	.773*	.056 (.029)	.150 (.044)	-.668	.808*	-.586*
update-3 & update-5	.172 (.052)	.065 (.041)	.591	.046 (.042)	.136 (.042)	-.634	.686**	-.454

* $p < .05$ ** $p < .01$

Highlights for Heinzel et al. Manuscript for NeuroImage (NIMG-15-3057R1)

- Working memory training leads to neural activation decrease in right MFG in aging
- Neural changes in transfer task are related to updating (compared to maintenance)
- Neural correlates of training and transfer effects overlap in right MFG
- Behavioral transfer to executive functions, processing speed, fluid intelligence
- Results advance understanding of neural mechanisms of training and transfer

52