

Study 1

Long-term brain effects of N-back training: an fMRI study

Anna Miró-Padilla, Elisenda Bueichekú, Noelia Ventura-Campos, Maria-Jesús Flores-Compañ, Maria-Antònia Parcet and César Ávila

Brain Imaging and Behavior

DOI: 10.1007/s11682-018-9925-x

ABSTRACT

Neurobehavioral effects of cognitive training have become a popular research issue. Specifically, behavioral studies have demonstrated the long-term efficacy of cognitive training of working memory functions, but the neural basis for this training have been studied only at short-term. Using fMRI, we investigate the cerebral changes produced by brief single *n*-back training immediately and 5 weeks after finishing the training. We used the data from a sample of fifty-two participants who were assigned to either an experimental condition (training group) or a no-contact control condition. Both groups completed three fMRI sessions with the same *n*-back task. Behavioral and brain effects were studied, comparing the conditions and sessions in both groups. Our results showed that *n*-back training improved performance in terms of accuracy and response speed in the trained group compared to the control group. These behavioral changes in trained participants were associated with decreased activation in various brain areas related to working memory, specifically the frontal superior/middle cortex, inferior parietal cortex, anterior cingulate cortex, and middle temporal cortex. Five weeks after training, the behavioral and brain changes remained stable. We conclude that cognitive training was associated with an improvement in behavioral performance and decreased brain activation, suggesting better neural efficiency that persists over time.

KEYWORDS

fMRI; *n*-back; cognitive training, long-term effects; working memory

1. Introduction

Working memory is necessary for a significant range of cognitive processes. It is important for everyday life because it is a determinant process in reasoning and in guiding decision-making and behavior (Diamond, 2013). In the past, working memory was defined as a rigid attribute, but it is now known that working memory can be improved when adequate training programs are used (Klingberg, 2010; Morrison & Chein, 2011; von Bastian & Oberauer, 2014). These behavioral studies have demonstrated both the immediate effects of this training and its long-term (2-12 months) efficacy once the training has ended. Cerebral changes produced by cognitive training have been studied, but only in the short-term, and so the stability over time of neural changes produced by this training remains unknown. Thus, the overall goal of the present study was to investigate the behavioral and cerebral changes produced by working memory training in the short and long-term.

The simple way of responding to stimuli and the easy management of difficulty makes the *n*-back task an appropriate tool to monitor working memory processes. A large number of studies have been conducted on cognitive training using *n*-back as the main task. All of them suggest that adequate *n*-back training improves task performance in terms of accuracy and reaction times, even with relatively short-term training (e.g. Buschkuhl et al., 2014; Hempel et al., 2004; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Küper & Karbach, 2016; Li et al., 2008; Salminen, Strobach, & Schubert, 2012; Schneiders, Opitz, Krick, & Mecklinger, 2011; Takeuchi et al., 2010; Thompson, Waskom, & Gabrieli, 2016; Yamashita, Kawato, & Imamizu, 2015). Generally, participants double or triple their pre-training performance levels (Kundu, Sutterer, Emrich, & Postle, 2013; Redick et al., 2013; Thompson et al., 2013; Jaeggi et al., 2008).

The researchers used training programs ranging from 60 to 1500 minutes, with no major differences in improvement (Anguera et al., 2012; Jaeggi et al., 2010b; Schneiders et al., 2012; Vartanian et al., 2013), but little research has focused on the effects of brief *n*-back training. The majority utilized an *n*-back adaptive task during training to manipulate the level of difficulty depending on the participant's performance. Both the single *n*-back and the dual *n*-back task have been used in training studies, with the latter being the most widely utilized, although both have shown efficacy in improving working memory capacity (Jaeggi et al., 2010b). In a very recent study, Küper & Karbach (2016) compared brief single *n*-back and dual *n*-back training, concluding that both showed equivalent improvement. Moreover, the authors concluded that in short periods of training, single *n*-back training can be more effective than dual *n*-back training (Küper & Karbach, 2016). Despite all the existing literature, only a few studies have tested the long-term (2-8 months) effects of *n*-back training, and they found that the behavioral changes observed remained stable (Jaeggi, Buschkuhl, Jonides, & Shah, 2011; Jaeggi, Buschkuhl, Shah, & Jonides, 2014; Katz et al., 2017; Li et al., 2008), although a decrease was observed in the performance between post-training and the follow-up session (Thompson et al., 2013).

N-back is one of the most common experimental paradigms for functional magnetic resonance imaging (fMRI) studies of working memory (Dobbs & Rule, 1989; Jaeggi et al., 2010a; Owen, McMillan, Laird, & Bullmore, 2005; Redick & Lindsey, 2013; Wager & Smith, 2003). In a meta-analysis by Owen et al. (2005), twenty-four fMRI studies with healthy subjects who performed the *n*-back task were analyzed in order to find the cerebral regions involved. They studied the brain areas activated depending on the type of stimulus used in the task. Their results showed six cortical regions and two subcortical

regions activated by verbal stimuli: the lateral premotor cortex, dorsal anterior cingulate and supplementary motor area, dorsolateral and ventrolateral prefrontal cortex, frontal pole, and bilateral and medial posterior parietal cortex; subcortically, the medial and lateral cerebellum and thalamus were activated (Owen et al., 2005).

There are few studies examining the cerebral changes produced by cognitive training on working memory (see the review by Buschkuhl, Jaeggy and Jonides, 2012). In that review, they concluded that there was evidence for brain changes in specific areas in terms of activation, but there was no agreement about whether this activation increased, decreased, or underwent redistribution, or even whether a reorganization of networks took place (Buschkuhl et al., 2012). First, regarding studies that reported increases in brain activation after working memory training, Buschkuhl et al. found limited evidence for this effect. For example, Westerberg & Klingberg (2007) evaluated the cerebral changes in only three young volunteers after five weeks of working memory training, and they found a significant increased activation in the middle or inferior frontal gyrus and parietal cortex. This finding coincides with the results from a similar prior experiment carried out by this group (Olesen, Westerberg, & Klingberg, 2004), where they found increased brain activity in the middle frontal gyrus and superior and inferior parietal cortices after five weeks of cognitive training. Second, regarding research that observed both increased and decreased activation (activation redistribution) after working memory training, one study by Dahlin et al. (2008; Experiment 1) stands out. In that study, analyses of pre- and post-training changes in the fMRI data showed increased activation in the left striatum, temporal, and occipital regions, but also decreased activity in frontal and parietal areas. Participants engaged in five weeks of computer-based updating training on a working

memory task. Finally, Buschkuhl et al. stated that no noteworthy working-memory training studies showed network reorganization.

Among fMRI studies that reported a decrease in cerebral activation, the majority used the *n*-back task for training (Hempel et al., 2004; Schneiders et al., 2011, 2012; Schweizer, Grahn, Hampshire, Mobbs, & Dalgleish, 2013; Thompson et al., 2016). Hempel et al. (2004) carried out four weeks of *n*-back training, and cerebral activation was examined before, after two weeks, and at the end of the training using fMRI. There was no control group. Their results showed an increased activation after two weeks of training in the right inferior frontal gyrus (BA 45) and the right intraparietal sulcus (BA 39/40), but activation decreased in the same areas after four weeks of *n*-back training, forming an inverse U-shaped activation pattern. In addition, Schneiders et al. (2011) used an 8-10 day adaptive *n*-back training (between 400-500 minutes) and performed fMRI pretest and posttest sessions. There were two training groups (visual or auditory *n*-back) and a no-contact control group. The authors observed decreased activation in the right superior middle frontal gyrus (BA 6/9/46) and right posterior parietal lobule (BA 40). In another study, Schneiders et al. (2012) reported the same activation pattern after training their participants for approximately the same length of time as in their prior study, but only on an adaptive auditory *n*-back task. Another study that used *n*-back for training was Schweizer et al. (2013), but in this case, the researchers trained their subjects on an adaptive affective *n*-back task for between 20-30 minutes during 20 days, and they had an active control group for comparison. They found activation decreases in the 3-back load level in the left dorsolateral prefrontal cortex, right superior frontal gyrus, bilateral supramarginal gyrus, bilateral middle temporal gyrus, and bilateral middle occipital lobe. In a recent fMRI study (Thompson et al. 2016) the participants were trained on an

adaptive dual *n*-back task in 20 sessions distributed across 4 weeks. Before and after the training, volunteers were scanned on the non-adaptive trained task. The authors used both an active and a no-contact control group. Once more, they reported reductions in activation in the bilateral inferior and middle frontal gyrus, insular cortex, and intraparietal sulcus. One study carried out by Vartanian et al. (2013) showed that very brief working memory training also causes cerebral modifications. They performed only 60 minutes of single *n*-back training and reported decreased brain activation in prefrontal areas (BA 46 and 47) (Vartanian et al., 2013). All the studies agree on the brain regions where the activation reductions were found.

On the other hand, Buschkuehl et al. (2014) wanted to test whether brief *n*-back training (less than 3h) increased task-related activation while participants performed difficult levels of the *n*-back task (4-back) using arterial spin labeling (ASL). They found that activation increased in prefrontal (BA6) and occipital (BA19) areas after training. 4-back places a high demand on cerebral resources, and it would involve an increment in the magnitude of perfusion (Buschkuehl et al., 2014). This activation pattern agrees with the compensation-related utilization of neural circuits hypothesis (CRUNCH) (Reuter-Lorenz & Cappell, 2008). This hypothesis postulates that people will activate more cortical regions as task load or resource demands increase. Previous results showed that at lower levels of task demands, older adults activate their task-related brain areas more than younger adults to achieve similar performance as younger adults. However, at harder levels, older people showed reduced task efficiency and less activation than young adults (Heinzel et al., 2016; Heinzel et al., 2014). Thus, in accordance with this hypothesis, the activations in task-related areas would increase or decrease depending on the difficulty of the task. Based on all of this literature, it is difficult to make predictions about cerebral

changes related to working memory training. Decreased activation is the most frequent result found after cognitive training in studies that use *n*-back, but it has been observed that with high-level demands, the activation increases (Buschkuhl et al., 2014).

In the present study, our main goal was to examine the long-term cerebral changes after working memory training because, to date, we did not find any study that held a follow-up session to evaluate the stability of the cerebral changes over time. By means of a longitudinal fMRI study, we examined the behavioral and functional data before a brief *n*-back training, immediately after it, and five weeks after finishing the training. Our participants trained for a total of 200 minutes on an adaptive version of single *n*-back on 1-back, 2-back and 3-back levels. That specific training was chosen because, based on previous findings mentioned above, in short periods of training, single *n*-back training would be more effective than dual *n*-back training. In the light of this, we hypothesized that: 1) Training processes would result in decreased activation of brain areas already involved in working memory, the frontal and parietal areas, in the short term; 2) The participants who trained on the adaptive *n*-back would produce faster responses and respond more accurately than non-trained participants immediately after *n*-back training, and this advantage would be maintained five weeks later; 3) After five weeks without training, the cerebral activation in the task-related brain areas would increase to compensate for the lack of training.

2. Materials and methods

2.1. Participants

Fifty-two healthy right-handed participants (21 male) with ages ranging between 21-26 years (mean age = 22.60±1.45) participated in this study. Subjects were recruited from

the student population of the University Jaume I, and none of them reported a previous psychiatric or neurologic diagnosis. Informed consent was obtained from each subject before participation, and they received monetary compensation for their time and effort. Participants were randomly assigned to either an experimental condition (training group) (N=25, mean age = 22.77 ± 1.5 , 9 men) or a control condition (control group) (N=27, mean age = 22.44 ± 1.4 , 12 men). Their intellectual level was assessed with the Matrix Reasoning Test (WAIS-III-R) (trained group: mean = 21.04 ± 3.42 ; control group: mean = 21.81 ± 2.02). The two groups differed only in the training (control group did nothing). The Ethical Committee of the Universitat Jaume I approved the research project.

2.2. Experimental paradigm

Both groups completed three fMRI sessions with the same adapted block-design *n*-back task (Zou et al., 2013). A schematic description of the experimental design is represented in **Figure 1**. The pre-training session, post-training session, and follow-up session correspond to Session 1 (S1), Session 2 (S2), and Session 3 (S3), respectively. Visual stimuli (letters) were presented electronically using E-Prime software (Psychology Software Tools, Pittsburgh, PA), professional version 2.0, installed in a Hewlett-Packard portable workstation (screen-resolution 800 x 600, refresh rate of 60 Hz). Participants watched the laptop screen through MRI-compatible goggles (VisuaStim, Resonance Technology, Inc., Northridge, CA, USA), and their responses were collected via MRI-compatible response-grips (NordicNeuroLab, Bergen, Norway). The E-Prime's logfile saved each participant's accuracy and reaction time (RTs) to each stimulus.

Fig. 1 Schematic description of the experimental design. Both groups performed the same n -back task in the three fMRI sessions. Training group data corresponds to the green lines and control group data to the blue lines. Training consisted in an adaptive n -back task, 4 sessions, 200 minutes



2.2.1. N -back fMRI task

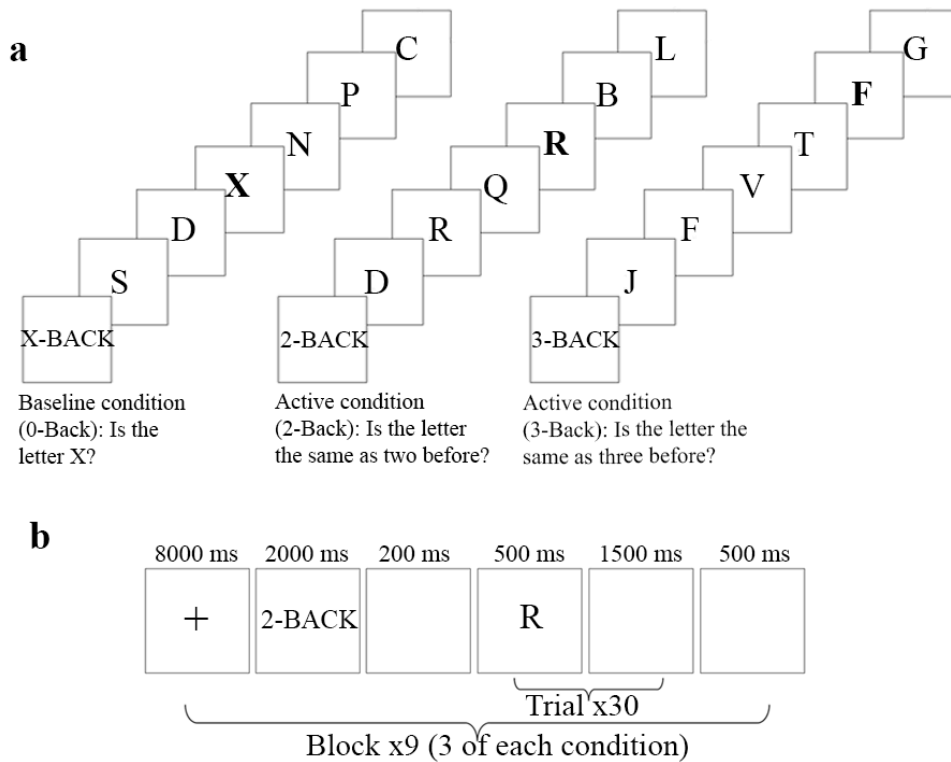
The task was presented in three load levels: two working memory blocks (2-back and 3-back) and a baseline control task (0-back). In 0-back, subjects pressed the “yes” button when the target (letter X) appeared on the screen, and they responded “no” to any other letters. In the 2-back and 3-back load levels, participants pressed the “yes” button when the current letter shown on the screen matched the one presented 2 or 3 items back, and they pressed “no” when there were no targets (see **Figure 2a**). Subjects had to give manual responses with only their right hand, responding to targets with their thumb and to non-targets with their forefinger.

With a total of nine blocks, three for each load level, the entire task lasted 11 minutes. Each block lasted 60.7 seconds and consisted of 200 ms of a blank screen, followed by 30 (6 target) consecutive trials of single letter stimuli (500 ms duration, 1500 ms inter-stimulus interval) with 500 ms of a blank screen at the end of each block. In addition, 8000 ms of a fixation cross and 2000 ms of an instruction display indicating task difficulty (0-back, 2-back or 3-back) were included before each block (see **Figure 2B**). There were 270 stimuli in all, and 54 of them were targets. The sequence of the stimuli was pseudo-randomized. The visual material comprised 15 different capital letters from the alphabet

(B, C, D, F, G, H, J, L, N, P, Q, R, S, T and V). Any letter could be a target in 2- and 3-back, but in 0-back only the “X” letter was the target. The letters, instructions, and fixation point were presented in the middle of the screen on a white background. All of them were in black ink with a 54-point *Arial* font. The task did not contain any lures.

Subjects received oral instructions about how to do the task, and they performed a 5-minute practicing task. In that, participants performed three blocks, one per load level, with only 15 trials (3 targets), in order to become familiar with the stimuli presentation and with the response buttons. A similar laptop with the same display features and the same hardware for manual responses was used outside of the scanner. Participants were asked to answer accurately and as quickly as possible.

Fig. 2 (a) Schematic example of the three load levels of our *n*-back fMRI task. Bold font indicates the correct response. (b) Block timing details of the *n*-back fMRI task



2.2.2. *N-back training task*

The training group carried out four consecutive sessions of single *n*-back training after fMRI S1 in our laboratory located at the University. One training session lasted 60 minutes and was distributed in two phases: the learning part and the test part. In the learning part, participants performed an adaptive *n*-back paradigm adapted from Jaeggi et al. (2008) for 50 minutes, whereas in the test part, they performed a simple *n*-back task, which lasted 10 minutes. Therefore, the total training time was approximately 200 minutes, plus 40 minutes for the test part. We used the same laptop as in the fMRI sessions, with the same display features and the same hardware for manual responses. Participants performed only one training session per day. As with our fMRI *n*-back task, no lures were present in our training task.

For the adaptive *n*-back task, we used the same stimuli and block timing as in the *n*-back fMRI task. However, we made some changes: the 0-back load level (0-back) disappeared, a new load level (1-back) was introduced, and participants were given feedback about their performance after each stimulus and at the end of each block. In 1-back, participants pressed the “yes” button when the current letter shown on the screen matched the one presented immediately before, and they pressed “no” in response to any other letters. We lengthened the task to approximately 16 minutes, and subjects performed three runs per training session. Once again, participants were asked to answer accurately and as quickly as possible.

In this task, we changed the level of difficulty by changing the level of “*n*” (1, 2 or 3) in order to motivate participants to improve. After each block, the participant’s individual performance was analyzed, and the *n*-back level was automatically adjusted. Thus, if the

participant had at least 90% correct answers, the level of “n” in the next block was increased by one, but it was decreased by one if accuracy was below 80%. In all other cases, the n-level remained constant (Salminen et al., 2012). In the last run, we increased the percentage by five percent to make it more difficult. Therefore, if the participant had at least 95% correct answers, the level of “n” was increased by one, whereas it was decreased by one if accuracy was below 85%. Each run started with the minimum level of “n” (1) for motivational reasons (Schneiders et al., 2012). Feedback was introduced after each response: a colored circle appeared for few seconds at the corner of the screen: green if the answer was correct, red if it was an error, and blue if participants did not press any button. Furthermore, at the end of each block, subjects received information about their performance: correct response percentage and reaction time average.

In the test part, participants performed an eight-block *n*-back task. We used the same stimuli and block timing as on the *n*-back fMRI task, but without the 0-back load level. Subjects had no feedback this time. Their results on this test were useful to evaluate their progress on *n*-back.

2.3. Neuroimaging data acquisition

Functional MRI data were collected on a 1.5T Siemens Symphony scanner (Erlangen, Germany). The same sequences were used in the three sessions. Participants were placed in a supine position in the MRI scanner, and their heads were immobilized with cushions to reduce motion artifacts. For task-fMRI, a gradient-echo T2*-weighted echo-planar MR sequence covering the entire brain was used (TR/TE=2500/49ms, matrix=64x64x28, flip angle = 90°, voxel size=3.5x3.5x4.48; slice thickness = 4 mm; slice gap = 0.48 mm). A total of 270 volumes were recorded. The slices were made parallel to the anterior–posterior commissure plane covering the entire brain. Before the functional magnetic

resonance sequences, a high-resolution structural T1-weighted MPRAGE sequence was acquired (TR = 2200 ms; TE = 3 ms; flip angle 90°, matrix = 256 x 256 x 160; voxel size = 1 x 1 x 1 mm).

2.4. Behavioral analysis

IBM SPSS Statistics software (Version 22 Armonk, New York, USA) was used to process the behavioral data (accuracy and RTs for participants' performance). A repeated-measures 2x3x3 mixed model ANOVA was conducted for each variable, using Group (training x control) as the between-subjects factor and Load Level (0-back vs. 2-back vs. 3-back) and Session (1 vs. 2 vs. 3) as within-subjects factors. Also, post-hoc analysis was conducted for each variable. With the test part data of the training, a repeated measures 2x4 ANOVA was conducted, with Load Level (2-back vs. 3-back) and Training Session (1 vs. 2 vs. 3 vs. 4) as within-subjects factors. Test-retest reliability analyses (re-test correlations) for behavioral control group data are provided in **Supplementary Information**.

2.5. Neuroimaging analysis

2.5.1. *Preprocessing*

Preprocessing and statistical analysis of fMRI data were conducted with SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK). We aligned each subject's fMRI data to the AC-PC plane by using his/her anatomical image. The fMRI preprocessing included head motion correction, where the functional images were realigned and resliced to fit the mean functional image. No participant had a head motion of more than 2.5 mm maximum displacement in any direction or 2.5° of any angular motion throughout the scan. Afterwards, the anatomical image (T1-weighted) was co-registered to the mean functional image, and the transformed anatomical image was then

re-segmented. The functional images were spatially normalized to the MNI (Montreal Neurological Institute, Montreal, Canada) space with 3 mm³ resolution, and spatially smoothed with an isotropic Gaussian kernel of 8 mm FWHM (Full-Width at Half-Maximum).

2.5.2. First level of analysis

Statistical analyses were performed in the context of the General Linear Model (Friston et al., 1995) for each participant and for each time point, using SPM12. In the first level analysis, we modeled the load levels of interest corresponding to 2-back > 0-back, 3-back > 0-back, and 2 and 3-back > 0-back. The BOLD signal was estimated by convolving the stimuli onset with the canonical hemodynamic response function. Six motion realignment parameters were included to explain signal variations due to head motion, that is, as covariates of no interest. A high-pass filter (128s) was applied to the functional data to eliminate low-frequency components. Then, contrast images were obtained to directly compare our load levels of interest. For the cross-sectional analysis, the first session (S1) load levels of interest were compared in order to assess differences between the *n*-back load levels before learning.

2.5.3. Statistical analysis

In the cross-sectional analysis, a whole-brain one-sample *t* test was conducted in order to study the brain regions involved in the *n*-back task (*2-back and 3-back load levels > 0-back load level*) using the fMRI data collected in S1. In addition, first session data were used to perform a two-sample *t* test to examine the equality of the brain responses in the two groups, so that between-groups brain differences found in subsequent sessions would be due to training effects. Test-retest reliability analyses (one sample *t* test) for imaging

control group data are provided in **Supplementary Information**.

In the second-level analysis, the longitudinal analysis was performed separately for 2-back and 3-back, with interaction analysis between sessions to evaluate: 1) the immediate effect of training, comparing S2 to S1; 2) the long-term effects of training, comparing S3 to S1; and 3) the effects of differences between immediate and long-term effects, comparing S3 to S2. To avoid false positives in the fMRI analyses (Woo, Krishnan, & Wager, 2014), the statistical criterion was set at $p < 0.05$, and family-wise error (FWE) was cluster-corrected for multiple comparisons (voxel-level uncorrected threshold of $p < 0.001$; specific cluster sizes appear in each result).

3. Results

3.1. Behavioral Results

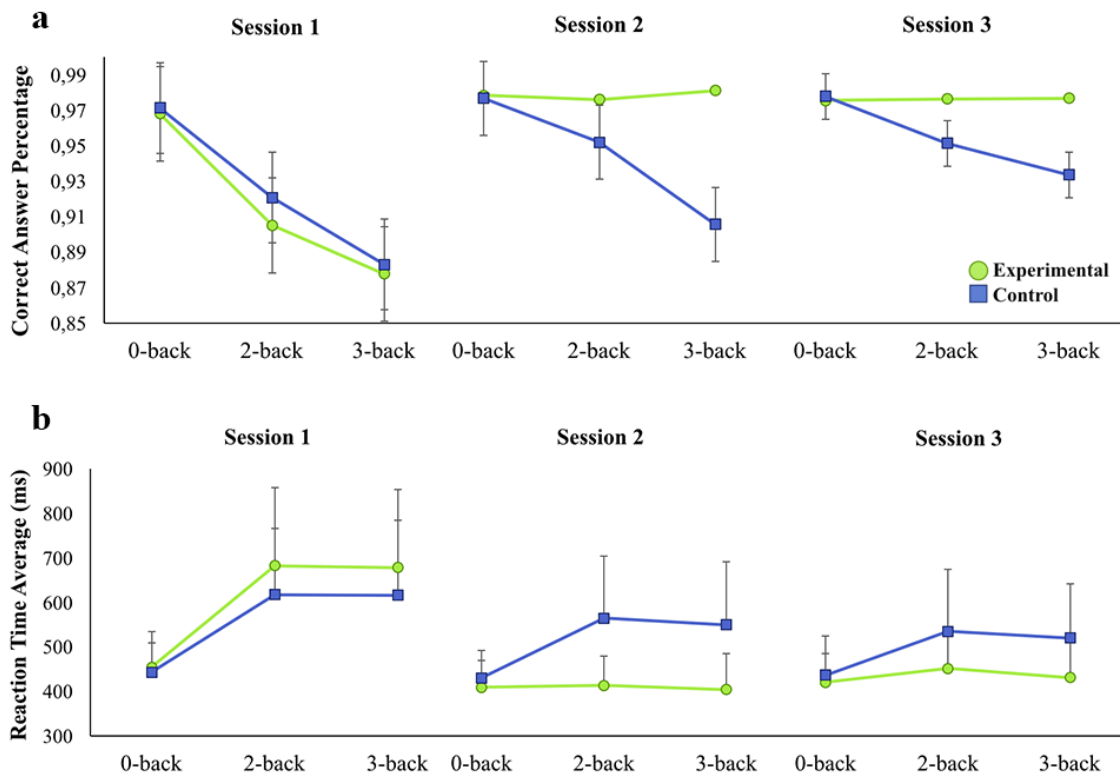
3.1.1 Behavioral fMRI results

The repeated measures $2 \times 3 \times 3$ mixed-model ANOVA conducted for accuracy yielded main effects for Session ($F_{(2,50)} = 34.66$ $p < .001$) and Load Level ($F_{(2,50)} = 42.85$ $p < .001$), which means that all the participants reduced their mistakes in the post-training and follow-up sessions, compared to S1, and that the highest accuracy scores were observed during the 0-back. These main effects were driven by significant Group x Session ($F_{(2,50)} = 7.77$ $p = .001$), Load Level x Session ($F_{(4,48)} = 13.07$ $p < .001$) and Load Level x Group ($F_{(2,50)} = 7.23$ $p = .002$) interactions. The first interaction indicated that trained participants were better than controls during the post-training and follow-up sessions, the second indicated that differences between load levels were greater at pre-training, whereas the third reflected that the training group showed better performance than the control group on 2-back and 3-back. As expected, the Load Level x Session x Group interaction reached significance ($F_{(4,48)} = 4.01$ $p = .007$), which means that the trained group became more

accurate in the post-training and follow-up sessions than the control group, when performing the 2-back and 3-back load levels (see **Figure 3a**). Post-hoc analyses revealed that these differences were significant for 3-back vs 0-back ($p=.002$), and they only approached significance for 2-back vs 0-back ($p=.13$).

Analyses of RTs scores revealed a similar pattern to that of accuracy. The 2x3x3 ANOVA also yielded significant main effects for Session ($F_{(2,50)} = 51.59$ $p<.001$) and Load Level ($F_{(2,50)} = 75.37$ $p<.001$). Both groups responded faster in the post-training and follow-up sessions than in the pre-training session. Also, both responded faster in the 0-back load level compared to the 2-back load level, as well as in the 2-back load level compared to the 3-back load level. Significant two-way interactions were obtained for the Group x Session ($F_{(2,50)}=28.14$ $p<.001$) and Load Level x Session ($F_{(4,48)}=28.23$ $p<.001$) interactions. The first two interactions may be interpreted similarly to accuracy, participants were faster than controls during the post-training and follow-up and the differences between load levels were greater at pre-training. Importantly, all these significant effects were qualified by the three-way Load Level x Session x Group interaction, which was highly significant ($F_{(4,48)} = 11.34$ $p<.001$). As expected, this interaction showed that the training group, compared to the controls, was faster after training and in the follow-up session in the 2-back and 3-back load levels (see **Figure 3b**). Post-hoc analyses revealed that this effect was significant for both 2-back vs 0-back and 3-back vs 0-back load levels ($p<.001$).

Fig. 3 Results of the behavioral analysis. (a) Correct-response percentage and (a) mean reaction times (in milliseconds) per session have been plotted as a function of load level and time. Pre-training session, post-training session and follow-up session correspond to Session 1, Session 2 and Session 3, respectively. Training group data correspond to the green lines (circles) and control group data to the blue lines (squares). RT = Reaction Time. Error bars represent standard error



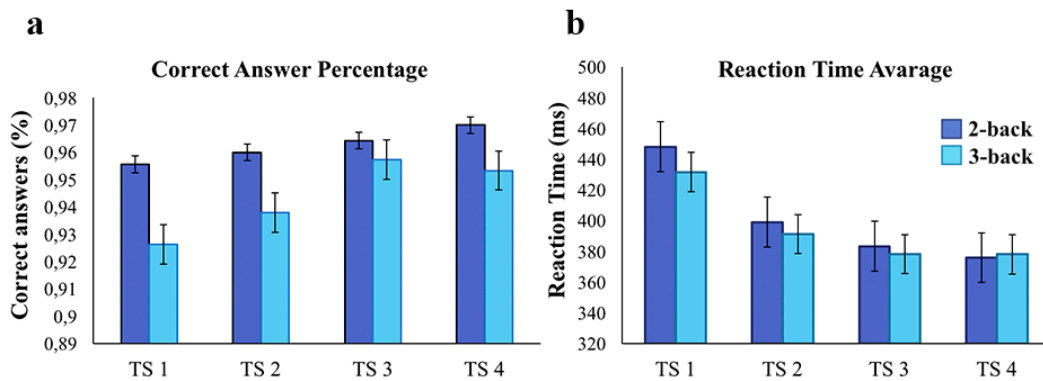
In sum, these results show that there were greater improvements in the 2-back and 3-back load levels after cognitive training, and that these improvements remained stable after five weeks.

3.1.2 Behavioral training results

With the behavioral training data for the training group, a repeated-measures 2x4 ANOVA was conducted with the results of the test part of the training to evaluate their progress on *n*-back. For accuracy training performance, a main effect of Training Session ($F_{(3,27)} = 6.49$ $p < .05$) and Load Level ($F_{(1,29)} = 11.99$ $p < .05$) was found, indicating participants' improvement, in terms of correct answers from one training session to

another, and reductions in their mistakes on both 2-back and 3-back. For RT values, we could see a significant effect of Training Session ($F_{(3,25)} = 10.35$ $p < .001$), which means that subjects' RTs decreased from one training session to another (see **Figure 4** for more values). As expected, these results confirmed the great progress of the training group on n -back performance after 200 minutes of training.

Fig. 4 Results of the behavioral training analysis of training group. (a) Correct-response percentage and (b) mean reaction times (in milliseconds) per training session have been plotted as a function of load level. 2-back data corresponds to the dark blue bars and 3-back data to the light blue bars. TS = training session RT = Reaction Time. Error bars represent standard error



3.2. Task fMRI Results

3.2.1 *Cross-Sectional analysis: Task effects at baseline*

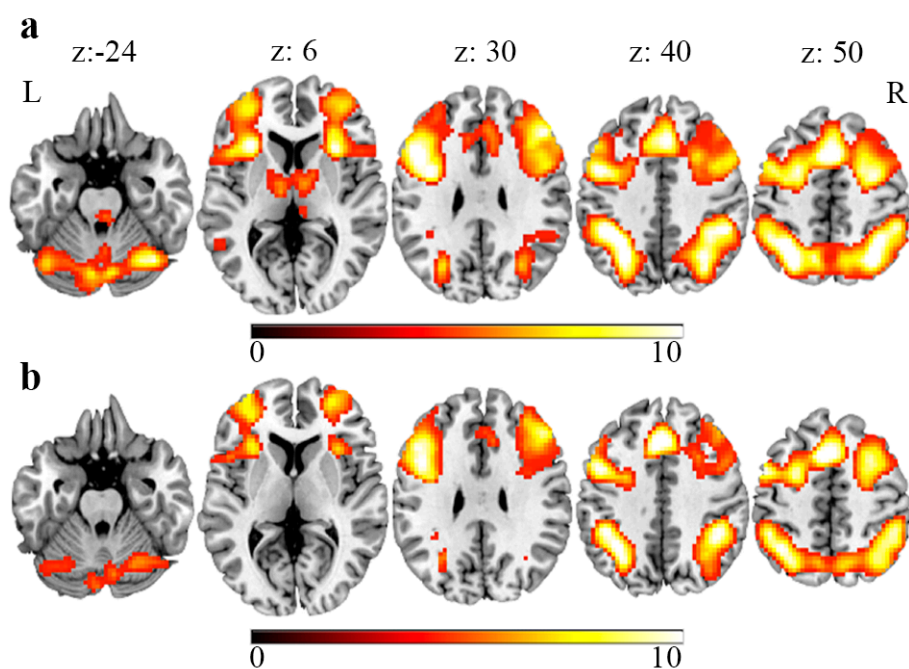
A whole-brain one-sample t test was conducted in order to study the brain regions involved in the n -back task (2-back and 3-back load levels $>$ 0-back load level). We used the fMRI data collected in S1. This analysis showed significant cortical and subcortical activations in brain areas related to working memory. Studying the task effects for each 2-back and 3-back load level (2-back $>$ 0-back and 3-back $>$ 0-back) separately (see **Figure 5**), the same areas were activated: bilateral superior, middle and inferior frontal cortex (BA 6/8-11/32/45-48), including supplementary motor area/anterior cingulate gyrus (SMA/ACC) (BA 6/32) and the insula (BA 47), bilateral superior and inferior parietal

cortex (BA 7/40), including precuneus, and bilateral cerebellum (crus I). Midbrain areas (thalamus and globus pallidus) were not significantly activated in 3-back, whereas they were in 2-back. Results were $p < .05$ FWE cluster-corrected using a threshold of $p < .001$ at the uncorrected voxel level with a cluster extension of $k = 2504$ voxels for 2-back and $k = 143$ for 3-back.

The two-sample t test analysis performed between groups to examine the equality in brain responses in S1 yielded no significant functional differences. As a result, the brain differences found between groups in subsequent sessions were due to training effects.

The threshold was $p < 0.001$ uncorrected at the voxel level.

Fig. 5 N -back general task activations. Both load levels presented an activation in the same areas, with some exception in cortical structures. (a) 2-back load level, represents the contrast: 2-back > 0-back and (b) 3-back load level, represents the contrast: 3-back > 0-back. Results were $p < .05$ FWE-cluster corrected with a cluster criteria of $k = 2504$ voxels for 2-back and $k = 143$ for 3-back. Left (L) and right (R). Coordinates are in the MNI space. Color bars express t -scores



3.2.2. Learning effects

To study the effects of training on the brain, an interaction analysis was conducted. Therefore, a 2x2 ANOVA (Group x Session) was carried out separately for each load level (2back and 3back). When studying the training effects by comparing S1 vs S2 and S1 vs S3 in the 2-back load level, we found similar results. These interaction analyses yielded activations in the bilateral superior frontal cortex (BA 8-9), including the SMA/ACC (BA 6/32), dorsolateral prefrontal cortex (BA 9/46), inferior frontal cortex (BA 44-46), and right inferior parietal cortex (IPC) (BA 39-40), in the trained group compared to the control group (see **Figure 6 and Table 1**). The reverse contrast yielded no significant effects. Results were $p < .05$ FWE cluster-corrected using a threshold of $p < .001$ at the uncorrected voxel level and a cluster extension of $k = 125$ voxels and $k = 87$ voxels, respectively.

Fig. 6 Results of the adaptive n -back post-training effects for 2-back load level: (a) represents the contrast: Trained group (S1>S2)>Control group (S1>S2) and (b) represents the contrast: Trained group (S1>S3)>Control group (S1>S3). Results were $p < .05$ FWE cluster-corrected using a threshold of $p < .001$ at the uncorrected voxel level and a cluster extension of $k = 125$ voxels and $k = 87$ voxels respectively. Left (L) and right (R). Coordinates are in the MNI space. Color bars express t -scores

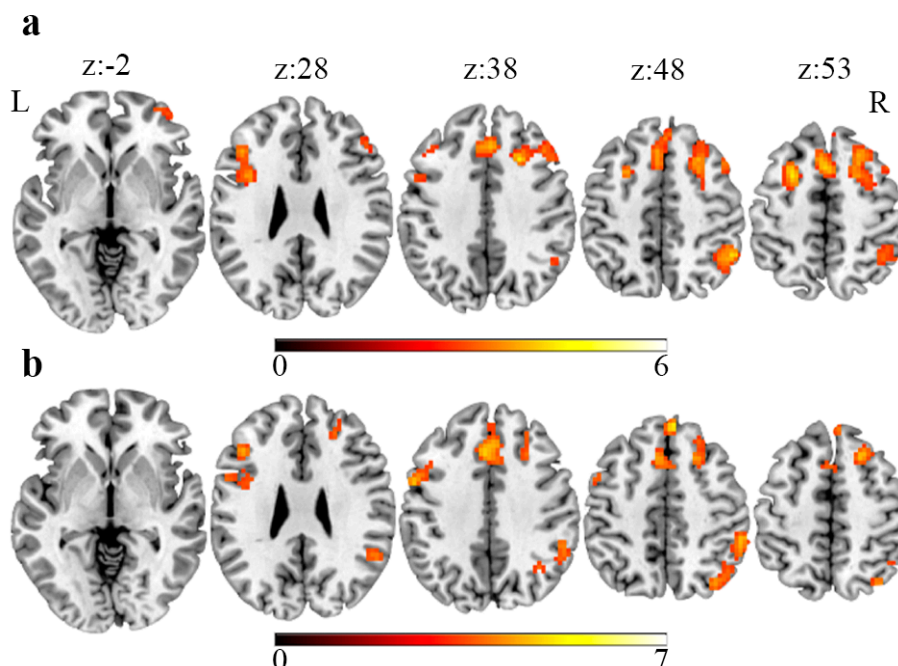


Table 1. List of brain activations as a result of the post-training session and follow-up session in 2-back load level between groups comparison: a) comparing Session 1 with Session 2 and b) comparing Session 1 with Session 3.

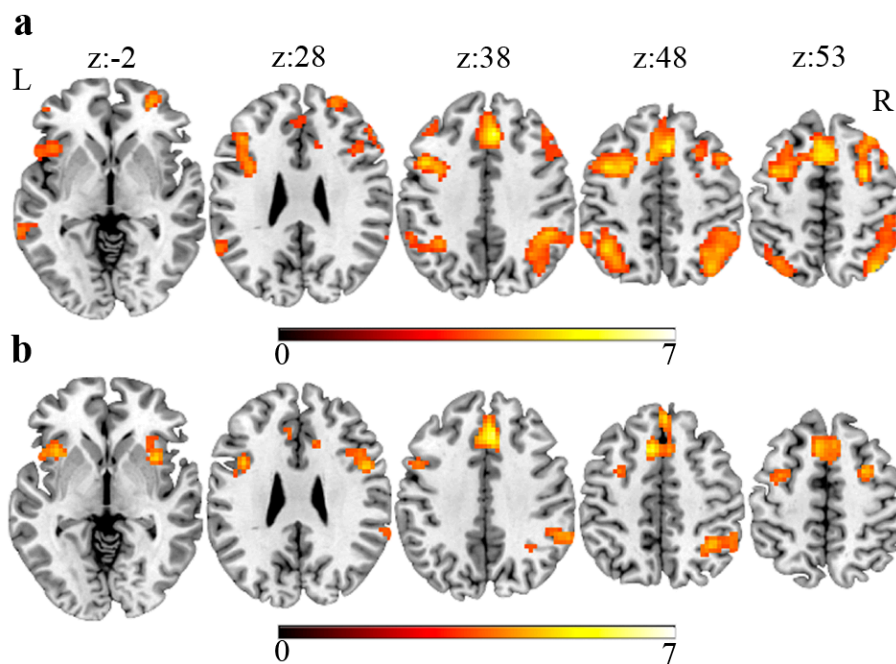
	BA	Cluster extent	MNI SPACE			Z-value
			x	y	z	
a) Trained Group (S1>S2) > Control Group (S1>S2)						
L Frontal Superior	6	132	-24	-1	59	4.67
R Frontal Middle	46	570	42	53	8	4.47
R Parietal Inferior	40	130	51	-49	47	4.21
SMA	6	248	0	14	53	4.16
L Frontal Inferior	48	125	-42	26	29	3.87
b) Trained Group (S1>S3) > Control Group (S1>S3)						
R Frontal Middle	9/8	235	3	44	44	4.64
R Parietal Inferior	40	121	54	-40	44	4.61
R Frontal Superior	8	141	24	26	59	4.53
L Frontal Middle	6	203	-54	2	38	4.24
R Parietal Superior	7	87	36	-70	50	4.04

FOOTNOTES: Results were $p < 0.05$ FWE cluster-corrected using a threshold of $p < 0.001$ at the uncorrected voxel level, and a cluster extension of $k = 125$ voxels and $k = 87$ voxels respectively. L = Left. R = Right. BA = Brodmann Area. SMA = supplementary motor area.

Regarding the 3-back load level, there were differences in the affected areas depending on the sessions compared. In the *Trained group (S1>S2) > Control group (S1>S2)* contrast, the analyses showed activations in the bilateral superior/middle frontal cortex (BA 8-11/46), including the SMA/ACC (BA 6/32), left insula, bilateral IPC (BA 39-40), and left temporal middle cortex (BA 21), in the trained group compared to the control group. On the other hand, in the *Trained group (S1>S3) > Control group (S1>S3)* contrast, the difference was found in the right IPC (BA 40), bilateral insula, SMA/ACC (6/32), bilateral inferior frontal cortex (BA 44-45), and dorsolateral prefrontal cortex (BA 9). The reverse contrasts yielded no significant differences. In **Figure 7 and Table 2**, we have included the results and values of these comparisons for each load level. The threshold was at $p < .05$ FWE cluster-corrected using an auxiliary threshold of $p < .001$ at

the uncorrected voxel level and a cluster extension of $k = 89$ voxels and $k = 67$ voxels, respectively.

Fig. 7 Results of the adaptive n -back post-training effects for 3-back load level: (a) represents the contrast: Trained group ($S1>S2$)>Control group ($S1>S2$) and (b) represents the contrast: Trained group ($S1>S3$)>Control group ($S1>S3$). Results were $p<.05$ FWE cluster-corrected using a threshold of $p<.001$ at the uncorrected voxel level and a cluster extension of $k = 89$ voxels and $k = 67$ voxels respectively. Left (L) and right (R). Coordinates are in the MNI space. Color bars express t -scores



When studying the stability of the effects of the working memory training over time, an interaction analysis was also conducted (*Trained group ($S2>S3$) > Control group ($S2>S3$)*) separately for 2-back and 3-back. No significant effects were found in either load level or any comparison. The threshold was $p<.001$ uncorrected at the voxel level.

In sum, comparing the pre-training session to the post-training session and the follow-up session, decreased activation was found in working memory brain areas when studying the 2-back or 3-back load level. These results were found for trained participants

compared to controls. However, no differences were found between the post-training and follow-up sessions, which means that the effects of training remained stable after 5 weeks.

Table 2. List of brain activations as a result of the post-training session and follow-up session in 3-back load level between groups comparison: a) comparing Session 1 with Session 2 and b) comparing Session 1 with Session 3.

	BA	Cluster extent	MNI SPACE			Z-value
			x	y	z	
a) Trained Group (S1>S2) > Control Group (S1>S2)						
R Frontal Superior	32/6	1380	6	29	41	5.48
L Frontal Middle	10	164	-36	53	11	4.78
R Frontal Middle	6	821	30	2	53	4.76
L Parietal Inferior	40	342	-39	-55	47	4.65
R Parietal Inferior	40	579	48	-43	44	4.65
L Temporal Middle	21	89	-66	-43	2	4.03
b) Trained Group (S1>S3) > Control Group (S1>S3)						
R Frontal Middle	32/9	454	6	29	38	5.08
L Insula	48/47	122	-33	17	-7	4.77
R Insula		110	36	11	-1	4.26
R Frontal Inferior	44	138	48	5	29	4.23
L Frontal Inferior	44	108	-39	8	29	4.16
R parietal Inferior	40	199	36	-52	47	4.09
L Frontal Middle	6	67	-33	1	53	3.74

FOOTNOTES: Results were $p < 0.05$ FWE cluster-corrected using a threshold of $p < 0.001$ at the uncorrected voxel level, and a cluster extension of $k = 89$ voxels and $k = 67$ voxels respectively. L = Left. R = Right. BA = Brodmann Area.

4. Discussion

The present fMRI research focused on studying the behavioral and neural changes associated with working memory training and their stability over time. To accomplish this, we randomly separated participants into two groups (training or control group), and both groups completed three fMRI sessions performing the same n -back task. The training group was trained outside of the scanner on an adaptive version of the single n -back task for 200 minutes in four training sessions between the pre-training and post-training sessions. A follow-up session was held after five weeks of no training. Our results

showed significant behavioral and functional differences between groups related to the working memory training. *N*-back training improved the performance on the task, and these behavioral changes were accompanied by decreased activation in diverse brain areas related to working memory, specifically, in the frontal superior/middle cortex, inferior parietal cortex, anterior cingulate cortex, and temporal middle cortex. Importantly, five weeks after the training, the behavioral and brain changes remained stable. Our results demonstrate that our cognitive training program improved behavioral performance and cause cerebral modifications that persist over time when compared with a no-contact control group.

Training effects were observed in terms of accuracy and RTs. Generally, all the participants in both groups improved their performance in the post-training session compared to the pre-training session. Control group improvements could be explained by retest effects due to task repetition, as reported in previous cognitive studies (Jaeggi et al., 2008; Schneiders et al., 2011). However, the training group reduced their errors and RTs significantly more than the control group in both working memory load levels. As expected, 200 minutes of working memory training on our adaptive single *n*-back task yielded an improvement in performance in terms of accuracy and reaction times. Our follow-up findings also showed that these behavioral changes remained stable five weeks after completing the training. A non-significant decrease was noted in the performance from the post-training session to the follow-up session because there was no additional training. These results agree with previous *n*-back training studies (e.g. Jaeggi et al., 2011; Thompson et al., 2013, 2016).

The *n*-back task activation pattern reported here, which includes the frontal, parietal, cerebellar, and subcortical areas, coincides with previous neuroimaging studies (Owen et al., 2005). All the activations found were bilateral and located specifically at the SMA/ACC (BA 6/32), superior, middle, and inferior frontal cortex, including the anterior insula, superior and inferior parietal cortex (BA7/40), including precuneus, cerebellum (crus I), and thalamus. Formerly, in working memory, the prefrontal cortex was considered a warehouse of information (Smith & Jonides, 1999), but current views give the prefrontal cortex the function of controlling the cognitive processing of information, selecting stimuli, and producing adequate responses (Postle, 2006). There is increasing evidence supporting this view (Lara & Wallis, 2015). In addition, executive manipulation of acquired facts has been associated with the parietal lobe (Koenigs, Barbey, Postle, & Grafman, 2009), as well as the storage function of working memory (Owen et al., 2005) as well as attentional processes of working memory (Berryhill, Chein, & Olson, 2011). Regarding the subcortical areas, the cerebellum assumes cognitive information processing functions due to its connections with the prefrontal cortex (Hayter, Langdon, & Ramnani, 2007; Vandervort, 2009). The thalamus, due to its attentional role of filtering relevant information, helps the prefrontal cortex in its working memory function (Watanabe & Funahashi, 2012).

Our findings are generally consistent with previous *n*-back functional neuroimaging studies that report decreased activation after training (Schneiders et al., 2011, 2012; Schweizer et al., 2013; Thompson et al., 2016). In relation to training activation changes, our imaging data revealed that participants who belonged to the training group showed decreased activation in various cerebral areas related to working memory. Decreased activation has been interpreted as an indication of better neural efficiency in these areas,

thus improving their function. This decline in cerebral activation may allow participants to respond more quickly and make fewer mistakes (Buschkuhl et al., 2014). Kelly, Foxe, & Garavan (2006) noted that this effect of decreased activation is typically observed after training on higher cognitive tasks, and they stated that lower activation is associated with increased neural efficiency, which means that fewer neurons are needed to give a fast and accurate answer to the task. However, some studies have criticized the better neural efficiency explanation for the decreases in activation for being overly simple and unclear (Constantinidis & Klingberg, 2016; Poldrack, 2015). In his review, Poldrack (2015) viewed efficiency as inverted energy for the transmission of information in the brain networks. He highlighted the need for new studies and models to examine the neural changes, and he reported that identifying potential activation effects may lead to future mechanistic explanations. Therefore, although a decrease in activation is often interpreted as an increase in neural efficiency in the literature, our data did not demonstrate the underlying cellular mechanism, but instead they pointed to the areas of change after working memory training. The bilateral superior frontal cortex (BA 8-9), IPC (BA 40) and SMA/ACC (BA 6/32) were the areas affected by this activation reduction in both 2-back and 3-back. During 3-back performance, we also found a decreased activation in the left insula and left middle temporal cortex (BA 21). The main effect only on 3-back performance in these specific areas may be due to more demanding load levels than those of 2-back (Thompson et al., 2016).

The activation decreases in the superior part of the frontal cortex in both hemispheres were expected because the dorsolateral prefrontal cortex is strongly involved in working memory processes (Lara & Wallis, 2015). It is essential for continuous updating processes, attention focus, and ordering and selecting stimuli, which are fundamental

processes in performing the *n*-back task successfully. Other areas related to working memory where this effect was found were the IPC, SMA and ACC. Regarding IPC, this area is in charge of the phonological store, as demonstrated in studies with patients with lesions in this area (Baldo & Dronkers, 2006). This storage of verbal information is necessary to carry out our *n*-back task because we used letters as stimuli. Moreover, the IPC is typically activated when an attentionally-demanding maintenance strategy is used (Berryhill et al., 2011). With regard to SMA, this area has been related to the planning of sequences of movement, motor learning, and motor activation of the hand. In our case, as participants had to give the answer by pressing a button with their right hand, the activation decreases in these areas were accompanied by a decrease in RTs. On the other hand, the ACC has been related to error detection (Bush, Luu, & Posner, 2000), which is crucial to carry out our working memory task. Menon & Uddin (2010) said that the ACC and the insula work together in the detection of important stimuli and in initiating attentional control signals. Regarding the middle temporal cortex, further investigation is needed to determine its exact relationship with working memory.

One of the novel goals of the present study was to investigate the long-term effects of cognitive training. We did not find any longitudinal fMRI research that studies the stability of brain changes produced by working memory training, and so we cannot compare our functional results. Our fMRI analysis showed no significant changes between the two sessions, which means that the changes due to *n*-back training remained stable after the training ended. Our findings showed that the main effect that occurred between the pre-training and post-training sessions (decrease in activation) was present in the same areas when comparing the pre-training session with the follow-up session. Thus, our results demonstrate that the behavioral and cerebral changes produced by

working memory training remain stable after five weeks without training. The stability of these brain changes after five weeks could suggest an improved efficiency of these areas because we found no modifications in the results when comparing S2 and S3 in the behavioral analysis or the fMRI analysis. The follow-up session seems to be a necessary component of any working memory training paradigm designed to create enduring improvements (Thompson et al., 2013).

Overall results are partially consistent with the CRUNCH theory (Reuter-Lorenz & Cappell, 2008). In fact, the reduction in activations after training in the training group may be explained by the theory because training reduced the required task demands. However, we also expected that 5 weeks of no training would increase the activations on the task, but this was not the case, indicating that the positive effects of training were maintained without any loss for at least five weeks, as the behavioral and neural data suggest.

This study has a few limitations. We used a no-contact control group that did not receive any training. The training group came to our laboratory on four consecutive days, and they had more contact with the experimenters than the control group did, which may result in motivational differences between the two groups in terms of task efficiency. Nonetheless, the control group improved their performance from S1 to S2 and from S2 to S3. Although this may be attributed to the re-test effect, we also note that this improvement would not have taken place if there had been a lack of motivation. In any case, in future studies, active control groups should be included in the study design because the observed gains may not be due to working memory training per se, but to the training in general. Another limitation may be the short training period (200 minutes),

although some studies have used the same training time or less and showed behavioral improvements and cerebral changes (Buschkuhl et al., 2014; Jaeggi et al., 2008; Küper & Karbach, 2016; Vartanian et al., 2013; Yamashita et al., 2015). In addition, we have chosen a brief single *n*-back training with an eye on future clinical interventions. A long training protocol might be difficult and costly for patients and institutions. Therefore, we wanted to evaluate the effects of this kind of short working memory training regime on healthy controls to allow comparisons with clinical populations in future studies. The practice is limited to 1-, 2- and 3-back levels, which may not seem challenging, but 3-back is considered a highly demanding task, and participants reported that they always tried to get better results because they could see their correct response percentage and reaction time average. Finally, the expression “long-term” should be used with care because five weeks is not much time compared to other studies investigating long-term effects for at least 3 months to 1 year (e.g. Jaeggi et al., 2014; Katz et al., 2017; Thompson et al., 2013). In any case, we think it is relevant that this is the first manuscript to investigate brain reorganization weeks after the training is over. Future studies should determine this stability in longer retest periods.

In conclusion, *n*-back training not only improves behavioral performance, but it also causes cerebral modifications as signaled by the decrease in the activation of various brain areas related to working memory. These behavioral and neural changes are stable and persist after weeks with no training on the task. The future challenge is to determine whether this kind of training has the same effects in a clinical population and could be translated into beneficial and long-lasting treatments, and test whether these changes last longer than five weeks.

Compliance with Ethical Standards:

Funding: This research was supported by grants from Jaume I University (P1•1B2013-63) and Ministerio de Economía y Competitividad (PSI2013-47504-R and PSI2016-78805-R). Authors AMP, EB and MJFC were supported by pre-doctoral graduate program grants (Jaume I University FPI to AMP; National FPU to EB; and National FPI to MJFC).

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional Review Board of the Universitat Jaume I and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

- Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuhl, M., Benson, B. L., Jennett, S., ... Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research*, 228(1), 107–115. <http://doi.org/10.1016/j.bbr.2011.11.040>
- Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, 20(5), 529–538. <http://doi.org/10.1037/0894-4105.20.5.529>
- Berryhill, M. E., Chein, J., & Olson, I. R. (2011). At the intersection of attention and memory: the mechanistic role of the posterior parietal lobe in working memory. *Neuropsychologia*, 49(5), 1306–15. <http://doi.org/10.1016/j.neuropsychologia.2011.02.033>
- Buschkuhl, M., Hernandez-Garcia, L., Jaeggi, S. M., Bernard, J. a., & Jonides, J. (2014). Neural effects of short-term training on working memory. *Cognitive, Affective & Behavioral Neuroscience*, 14(1), 147–60. <http://doi.org/10.3758/s13415-013-0244-9>
- Buschkuhl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience*, 2, Supplem, S167–S179. <http://doi.org/10.1016/j.dcn.2011.10.001>
- Bush, G., Luu, P., & Posner, M. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.*, 4(6), 215–222. [http://doi.org/10.1016/S1364-6613\(00\)01483-2](http://doi.org/10.1016/S1364-6613(00)01483-2)
- Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, 17(7), 438–449. <http://doi.org/10.1038/nrn.2016.43>
- Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of Learning After Updating Training Mediated by the Striatum. *Science*, 320(5882),

- 1510–1512. <http://doi.org/10.1126/science.1155466>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–68. <http://doi.org/10.1146/annurev-psych-113011-143750>
- Dobbs, a R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging*, 4(4), 500–503. <http://doi.org/10.1037//0882-7974.4.4.500>
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical Parametric Maps in Functional Imaging : A General Linear Approach. *Human Brain Mapping*, 2(4), 189–210. <http://doi.org/10.1002/hbm.460020402>
- Hayter, A. L., Langdon, D. W., & Ramnani, N. (2007). Cerebellar contributions to working memory. *NeuroImage*, 36(3), 943–954. <http://doi.org/10.1016/j.neuroimage.2007.03.011>
- Heinzel, S., Lorenz, R. C., Brockhaus, W.-R., Wüstenberg, T., Kathmann, N., Heinz, A., & Rapp, M. A. (2014). Working Memory Load-Dependent Brain Response Predicts Behavioral Training Gains in Older Adults. *The Journal of Neuroscience*, 34(4), 1224–1233. <http://doi.org/10.1523/JNEUROSCI.2463-13.2014>
- Heinzel, S., Lorenz, R. C., Pelz, P., Heinz, A., Walter, H., Kathmann, N., ... Stelzel, C. (2016). Neural correlates of training and transfer effects in working memory in older adults. *NeuroImage*, 134, 236–249. Retrieved from <http://dx.doi.org/10.1016/j.neuroimage.2016.03.068> \npapers://70c55321-b126-46ea-ac63-5b3363fb7a4a/Paper/p2623
- Hempel, A., Giesel, F. L., Garcia Caraballo, N. M., Amann, M., Meyer, H., Wüstenberg, T., ... Schröder, J. (2004). Plasticity of Cortical Activation Related to Working Memory during Training. *American Journal of Psychiatry*, (April), 745–747. <http://doi.org/10.1176/appi.ajp.161.4.745>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), 6829–6833. <http://doi.org/10.1073/pnas.0801268105>
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010a). The concurrent validity of the N -back task as a working memory measure. *Memory*, 18(4), 394–412. <http://doi.org/10.1080/09658211003702171>
- Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010b). The relationship between n-back performance and matrix reasoning — implications for training and transfer. *Intelligence*, 38(6), 625–635. <http://doi.org/10.1016/j.intell.2010.09.001>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108(25), 10081–10086. <http://doi.org/10.1073/pnas.1103228108>
- Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, 42, 464–80. <http://doi.org/10.3758/s13421-013-0364-z>
- Katz, B., Au, J., Buschkuhl, M., Abagis, T., Zabel, C., Jaeggi, S. M., & Jonides, J. (2017). Individual Differences and Long-term Consequences of tDCS-augmented Cognitive Training. *Journal of Cognitive Neuroscience*, 29(9), 1498–1508. http://doi.org/10.1162/jocn_a_01115
- Kelly, C., Foxe, J. J., & Garavan, H. (2006). Patterns of Normal Human Brain Plasticity After Practice and Their Implications for Neurorehabilitation. *Archives of Physical Medicine and Rehabilitation*, 87(12 SUPPL.), 20–29. <http://doi.org/10.1016/j.apmr.2006.08.333>

- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317–324. <http://doi.org/10.1016/j.tics.2010.05.002>
- Koenigs, M., Barbey, A. K., Postle, B. R., & Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 29(47), 14980–6. <http://doi.org/10.1523/JNEUROSCI.3706-09.2009>
- Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(20), 8705–15. <http://doi.org/10.1523/JNEUROSCI.5565-12.2013>
- Küper, K., & Karbach, J. (2016). Increased training complexity reduces the effectiveness of brief working memory training: evidence from short-term single and dual n-back training interventions. *Journal of Cognitive Psychology*, 28(2), 199–208. <http://doi.org/10.1080/20445911.2015.1118106>
- Lara, A. H., & Wallis, J. D. (2015). The role of prefrontal cortex in working memory: a mini review. *Frontiers in Systems Neuroscience*, 9(December), 1–7. <http://doi.org/10.3389/fnsys.2015.00173>
- 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. *Psychology and Aging*, 23(4), 731–742. <http://doi.org/10.1037/a0014343>
- Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, 214(5-6), 655–667. <http://doi.org/10.1007/s00429-010-0262-0>
- Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, 18(1), 46–60. <http://doi.org/10.3758/s13423-010-0034-0>
- Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, 7(1), 75–9. <http://doi.org/10.1038/nn1165>
- 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. *Human Brain Mapping*, 25(1), 46–59. <http://doi.org/10.1002/hbm.20131>
- Poldrack, R. A. (2015). Developmental Cognitive Neuroscience Is “efficiency” a useful concept in cognitive neuroscience? *Developmental Cognitive Neuroscience*, 11, 12–17. <http://doi.org/10.1016/j.dcn.2014.06.001>
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, 139(1), 23–38. <http://doi.org/10.1016/j.neuroscience.2005.06.005>
- Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 20(6), 1102–1113. <http://doi.org/10.3758/s13423-013-0453-9>
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... Engle, R. W. (2013). No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study. *Journal of Experimental Psychology: General*, 142(2), 359–379. <http://doi.org/10.1037/a0029082>
- Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive Aging and the Compensation Hypothesis. *Current Directions in Psychological Science*, 17(3), 177–182. <http://doi.org/10.1111/j.1467-8721.2008.00570.x>

- Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working memory training on executive functioning. *Frontiers in Human Neuroscience*, 6(June), 1–14. <http://doi.org/10.3389/fnhum.2012.00166>
- Schneiders, J. a, Opitz, B., Krick, C. M., & Mecklinger, A. (2011). Separating intra-modal and across-modal training effects in visual working memory: an fMRI investigation. *Cerebral Cortex (New York, N.Y.: 1991)*, 21(11), 2555–64. <http://doi.org/10.1093/cercor/bhr037>
- Schneiders, J., Opitz, B., Tang, H., Deng, Y., Xie, C., Li, H., & Mecklinger, A. (2012). The impact of auditory working memory training on the fronto-parietal working memory network. *Frontiers in Human Neuroscience*, 6(June), 173. <http://doi.org/10.3389/fnhum.2012.00173>
- Schweizer, S., Grahn, J., Hampshire, A., Mobbs, D., & Dalgleish, T. (2013). Training the emotional brain: Improving affective control through emotional working memory training. *Annals of Internal Medicine*, 158(6), 5301–5311. <http://doi.org/10.1523/JNEUROSCI.2593-12.2013>
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science (New York, N.Y.)*, 283(5408), 1657–61. <http://doi.org/http://dx.doi.org/10.1016/B978-0-12-385157-4.01148-9>
- Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., ... Kawashima, R. (2010). Training of Working Memory Impacts Structural Connectivity. *Journal of Neuroscience*, 30(9), 3297–3303. <http://doi.org/10.1523/JNEUROSCI.4611-09.2010>
- Thompson, T. W., Waskom, M. L., & Gabrieli, J. D. E. (2016). Intensive Working Memory Training Produces Functional Changes in Large-scale Frontoparietal Networks. *Journal of Cognitive Neuroscience*, 28(4), 575–88. http://doi.org/10.1162/jocn_a_00916
- Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., ... Gabrieli, J. D. E. (2013). Failure of Working Memory Training to Enhance Cognition or Intelligence. *PLoS ONE*, 8(5). <http://doi.org/10.1371/journal.pone.0063614>
- Vandervert, L. (2009). WorkingMemory, the Cognitive Functions of the Cerebellum and the Child Prodigy. In: L. Shavinina, ed., *International Handbook on Giftedness*, 1st ed. pp.3-13. Springer Science+Business Media B.V. doi: 10.1007/978-1-4020-6162-2 13
- Vartanian, O., Jobidon, M.-E., Bouak, F., Nakashima, A., Smith, I., Lam, Q., & Cheung, B. (2013). Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. *Neuroscience*, 236, 186–194. <http://doi.org/10.1016/j.neuroscience.2012.12.060>
- von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research*, 78(6), 803–820. <http://doi.org/10.1007/s00426-013-0524-6>
- Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective and Behavioral Neuroscience*, 3(4), 255–274. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-2342597739&partnerID=40&md5=d13bba90ed9b65ae16fd931accf885b1>
- Watanabe, Y., & Funahashi, S. (2012). Thalamic mediodorsal nucleus and working memory. *Neuroscience and Biobehavioral Reviews*, 36(1), 134–42. <http://doi.org/10.1016/j.neubiorev.2011.05.003>
- Westerberg, H., & Klingberg, T. (2007). Changes in cortical activity after training of working memory - a single-subject analysis. *Physiology and Behavior*, 92(1-2),

- 186–192. <http://doi.org/10.1016/j.physbeh.2007.05.041>
- Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *NeuroImage*, 91, 412–9. <http://doi.org/10.1016/j.neuroimage.2013.12.058>
- Yamashita, M., Kawato, M., & Imamizu, H. (2015). Predicting learning plateau of working memory from whole-brain intrinsic network connectivity patterns. *Scientific Reports*, 5, 7622. <http://doi.org/10.1038/srep07622>
- Zou, Q., Ross, T. J., Gu, H., Geng, X., Zuo, X.-N., Hong, L. E., ... Yang, Y. (2013). Intrinsic resting-state activity predicts working memory brain activation and behavioral performance. *Human Brain Mapping*, 34(12), 3204–15. <http://doi.org/10.1002/hbm.22136>

SUPPORTING INFORMATION

Test-retest reliability analyses for behavioral and imaging control group data:

Behavioral test-retest reliability analyses

We calculated the re-test correlations separately for accuracy and reaction times (RTs). The results yielded significant positive correlations: Session 1 - Session 2: $r_{27} = 0.53$ $p < 0.05$; Session 1 – Session 3: $r_{27} = 0.54$ $p < 0.05$ and Session 2 – Session 3: $r_{27} = 0.71$ $p < 0.01$ for accuracy and Session 1 - Session 2: $r_{27} = 0.88$ $p < 0.01$; Session 1 – Session 3: $r_{27} = 0.86$ $p < 0.01$ and Session 2 – Session 3: $r_{27} = 0.95$ $p < 0.01$ for RTs.

Imaging test-retest reliability analyses

For imaging data, we performed three one sample t tests (one per session) in order to observe the reliability. Here, we present the results in a figure:

Supplementary Figure: One Sample t test for the control group in (a) Session 1, (b) Session 2, and (c) Session 3, showing the n -back general activations. The same contrast was utilized for the 3 sessions: 2-back+3-back>0-back. Results were $p<.05$ FWE-cluster corrected with cluster criteria of $k = 657$ voxels for Session 1, $k = 101$ for Session 2, and $k = 161$ voxels for Session 3. L = left and R = right.

