Cerebral Cortex, May 2017;27:2911–2925

doi:10.1093/cercor/bhw141
Advance Access Publication Date: 25 May 2016

ORIGINAL ARTICLE

Repeated Structural Imaging Reveals Nonlinear Progression of Experience-Dependent Volume Changes in Human Motor Cortex

Elisabeth Wenger[1], Simone Kühn[1,2], Julius Verrel[1], Johan Mårtensson[3], Nils Christian Bodammer[1], Ulman Lindenberger[1,4] and Martin Lövdén[1,5]

[1]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany, [2]Clinic and Policlinic for Psychiatry and Psychotherapy, University Clinic Hamburg-Eppendorf, Hamburg, Germany, [3]Department of Psychology, Lund University, Lund, Sweden, [4]European University Institute, San Domenico di Fiesole (FI), Italy and [5]Aging Research Center, Karolinska Institutet and Stockholm University, Stockholm, Sweden

Address correspondence to Elisabeth Wenger, Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Email: wenger@mpib-berlin.mpg.de

Abstract

Evidence for experience-dependent structural brain change in adult humans is accumulating. However, its time course is not well understood, as intervention studies typically consist of only 2 imaging sessions (before vs. after training). We acquired up to 18 structural magnetic resonance images over a 7-week period while 15 right-handed participants practiced left-hand writing and drawing. After 4 weeks, we observed increases in gray matter of both left and right primary motor cortices relative to a control group; 3 weeks later, these differences were no longer reliable. Time-series analyses revealed that gray matter in the primary motor cortices expanded during the first 4 weeks and then partially renormalized, in particular in the right hemisphere, despite continued practice and increasing task proficiency. Similar patterns of expansion followed by partial renormalization are also found in synaptogenesis, cortical map plasticity, and maturation, and may qualify as a general principle of structural plasticity. Research on human brain plasticity needs to encompass more than 2 measurement occasions to capture expansion and potential renormalization processes over time.

Key words: gray matter changes, motor learning, structural brain plasticity, time course, voxel-based morphometry

Introduction

Following Lövdén et al. (2010, 2013), we define plasticity as the inherent ability of the brain to undergo macroscopic structural change in response to altered environmental demands. In contrast, we define flexibility as the adaptive reconfiguration of the existing behavioral repertoire in the absence of macroscopic structural change. Plastic changes are triggered in the presence of a prolonged mismatch between the functional supply of brain structure and the experiential demands of the environment that cannot be accommodated by flexibility. Plastic changes thus require but are not synonymous with plasticity itself, that is, the potential for plastic change.

There is accumulating evidence for such experience-dependent plastic changes in the structure of the adult brain (for review, see Hübener and Bonhoeffer 2014), including macrostructural changes in the brains of adult humans (e.g., Draganski et al. 2004; Woollett and Maguire 2011). Using structural, T_1-weighted magnetic resonance (MR) imaging, gray matter alterations have been observed following extensive behavioral interventions,

such as several months of juggling training (Draganski et al. 2004), intensive studies for medical exams (Draganski et al. 2006), foreign language studies (Mårtensson et al. 2012), spatial navigation training (Lövdén et al. 2012; Wenger et al. 2012), and video game playing (Kühn et al. 2014). Other studies have reported gray matter changes after 2 weeks of mirror reading (Ilg et al. 2008), 7 days of juggling training (Driemeyer et al. 2008), a few days of signature writing (Hamzei et al. 2012), and even after only 2 sessions of practice in a complex whole-body balancing task (Taubert et al. 2010) or hours of training on color subcategories (Kwok et al. 2011). This suggests that changes in gray matter volume may emerge quite rapidly.

Several questions related to experience-dependent changes in human gray matter structure remain unanswered. For example, the associations between neural and behavioral changes remain unclear. It is tempting to think that changes in brain volume are direct manifestations of task proficiency and thus can be regarded as neural correlates of skill acquisition and knowledge accumulation. However, volume changes may also be induced by regional alterations in neural activity that index time on task or effort invested regardless of learning success (Lövdén et al. 2013; Bellander et al. 2016). Also, plastic changes in human brain structure have typically been observed with pretest–posttest designs incorporating only 2 measurement time points with structural imaging data. These studies implicitly assume, by virtue of their design, that learning is accompanied by a monotonic increase of gray matter structure, such as continuous linear or asymptotic increase throughout the time of training. In contrast, results from some animal studies would suggest an inverse quadratic shape, that is, an initial expansion followed by (partial) renormalization. For example, in vivo microscopic imaging of dendritic spines in mice revealed new spines after a few hours of motor training (Xu et al. 2009). These rapid changes were followed by selective stabilization of new spines while older spines were partly eliminated, which partially renormalized overall spine density (see Fu and Zuo 2011 for review). Learning-related cortical map expansion has also been shown to occur quite rapidly (i.e., within a few days) and then renormalize during further training despite stable performance (Molina-Luna et al. 2008; Reed et al. 2011). It has been proposed that an initial "overshoot" may increase the pool of neural resources from which the most efficient wiring can then be selected (Reed et al. 2011). Quallo et al. (2009) analyzed structural data of 3 adult macaque monkeys, collected on multiple occasions before, during, and after learning to use a rake for retrieving food. They found learning-related increases in task-relevant brain regions, which also mapped onto the learning curves. Crucially, despite continued training, the observed increased gray matter structure decreased again after the monkey's performance reached asymptote. After training, the volume was still enlarged compared with before training, but much smaller in magnitude than the peak effect observed before asymptotic performance was reached.

Taken together, these results from animal literature and the reports in humans of structural alterations following very different training periods (e.g., Draganski et al. 2004; Ilg et al. 2008; Kwok et al. 2011; Woollett and Maguire 2011) call for a closer investigation of the temporal dynamics of gray matter changes. While a few studies have started to address the question of time scale of experience-dependent structural changes in humans (Driemeyer et al. 2008; Taubert et al. 2010; Hamzei et al. 2012), conclusive studies on the shape of gray matter changes over time are still needed (May 2011; Lövdén et al. 2013).

Numerous studies showing neuroplasticity investigated different types of motor skills, such as typing, juggling, or playing an instrument, to name a few. Motor skill learning refers to the process by which movements are executed more quickly and accurately with practice (Willingham 1998). Hikosaka et al. (2002) suggested a model of staged motor learning that distinguishes between 2 types of information being processed separately and most likely consecutively: a spatial processing stream that encodes the visuospatial coordinates of the newly learnt movement (involving the basal ganglia, prefrontal and parietal cortices, and cerebellum) and a movement processing stream that encodes the motor program that initiates the corresponding muscle activity (involving again basal ganglia, motor cortex, and cerebellum; Hikosaka et al. 2002). The focus on motor skills has several benefits: It serves as a bridge to animal research, comes with clear expectations about relevant brain regions, and provides reliable and valid assessments of progress in the target of training. Hence, motor skill learning has become a preferred vehicle for investigating human neuroplasticity.

To test the shape of gray matter changes over time and the divergent predictions of net growth versus initial expansion followed by renormalization, we trained 15 right-handed men (aged 25–36) to write and draw with their nondominant (i.e., left) hand on a tablet computer for 30–45 min per day for a period of 7 weeks. During this 7-week period, we acquired functional and structural magnetic resonance images at up to 18 occasions. Sixteen right-handed, age-matched participants were measured as a control group before the experimental group's training period (i.e., pretest), one time in the middle, (i.e., in Week 4), and after the training period (i.e., posttest). We chose this specific motor paradigm based on previous studies reporting changes in gray matter structure following the acquisition of fine motor skills in humans (Granert et al. 2011; Gryga et al. 2012; Hamzei et al. 2012) and considering that brain regions underlying motor learning are relatively well researched (see above, e.g., Doyon and Benali 2005; Luft and Buitrago 2005). We mainly expected effects in the areas of primary motor cortices (i.e., the precentral gyri) that are involved in hand motor action, given their important role in acquiring fine motor skills (Sanes and Donoghue 2000; Ungerleider et al. 2002). Effects in the postcentral gyri, basal ganglia, and cerebellum may also be expected (Katanoda et al. 2001; Harrington et al. 2007; Doyon et al. 2009).

Materials and Methods

Participants

Thirty-one male adults between 25 and 36 years (M_{age} = 28.48, SD = 2.45) were recruited through flyers, word-of-mouth recommendation, or following their participation in previous studies. All of them were strongly right-handed (value of >0.80 in the Edinburgh Handedness Inventory; Oldfield 1971), had normal or corrected-to-normal vision, had no history of psychological or neurological diseases, and did not have any contraindication to participate in an MR study, like metallic implants, tattoos, tinnitus, or claustrophobia. None of the participants were professional musicians or practiced or had practiced a musical instrument on a daily basis. None reported any injuries of their hands or arms in the past 3 years. After pretest, participants were randomly assigned to either experimental (n = 15, M_{age} = 28.53, SD = 2.39, $M_{Edinburgh\ score}$ = 0.89, SD = 0.09) or control group (n = 16, M_{age} = 28.44, SD = 2.58, $M_{Edinburgh\ score}$ = 0.88, SD = 0.09). The groups did not differ with respect to age ($t_{(29)} < 1$, $P = 0.92$) or degree of handedness ($t_{(29)} < 1$, $P = 0.60$).

Participants in the experimental group were paid up to 1000€, and participants in the control group were paid up to 580€ for completion of the whole study. The ethical board of the DGPs (Ethikkommission der Deutschen Gesellschaft für Psychologie) approved the study, and written informed consent of all participants was obtained prior to the investigation.

General Study Design

All participants underwent MR measurement and extensive behavioral testing at pretest. After this, participants of the experimental group started the training routine of the left hand, consisting of writing and drawing on a touch-sensitive tablet PC for 30 to 45 min per day for 7 weeks (see below for details). In addition, they were invited to our MR lab 3 times per week in the first 2 weeks of training and twice per week in the remaining 5 weeks of training, to undergo both structural and functional scanning. Participants of the control group followed their normal daily routines without any intervention and were scanned once in Week 4 of the experimental group's training period. After the training period, both the experimental and control group were invited again to complete another behavioral assessment and MR measurement for posttest, involving the same tasks assessed at pretest. This study design resulted in a total number of 18 MR measurements per person in the experimental group and 3 MR measurements per person in the control group.

Pretest and Posttest Measurements

In a behavioral session, participants were asked to complete a short demographic questionnaire and the Edinburgh Handedness Inventory (Oldfield 1971). In a scanner simulator session, participants were trained to feel comfortable with the MRI environment and familiarized with the functional MRI (fMRI) tasks that were later administered during the first imaging session. The MR measurement protocol for pre- and posttest was measured on 2 consecutive days and included a T_1-weighted structural scan, fMRI, and DTI (see MR Image acquisition for sequence parameters). To exclude potential short-term, activity-related effects on brain structure, participants were instructed not to perform their daily training session before coming to the MR laboratory, but only afterwards.

To identify regions functionally relevant for the execution of the trained left-hand writing and drawing task, we administered a "writing task" inside the scanner (first fMRI task). Participants had a pillow on their lap, on top of which they had an A5-sized notebook and a pencil. The head coil was equipped with 2 mirrors such that participants could see their own hands with the book and pencil. Red, green, or yellow light was projected into the scanner tunnel: whenever green was presented, participants were asked to write the letter sequence "lalala," whereas they were to relax their hand when there was red light ("rest" condition). During the first block of the task, participants wrote with their right hand. During the second block, indicated by yellow light, they wrote with their left hand. Without any behavioral endpoint measures, this task mainly served as a localizer task for the regions functionally involved in left- and right-hand writing movements (Katanoda et al. 2001; Harrington et al. 2007; Tam et al. 2011). The second fMRI task ("tapping task") required participants to tap their thumb and index finger against one another at a rate of about 1 Hz. They did this with their right hand when they saw the letter R on the screen and with their left hand when they saw the letter L on the screen. This task served as a localizer for the cortical map representation of the respective hand (Gelnar et al. 1999). Finally, we performed diffusion-tensor imaging (DTI) to get mean diffusivity (MD) and fractional anisotropy (FA) values for gray matter regions showing structural change. The MD values were later used as a proxy for tissue density in gray matter.

Training Paradigm

The behavioral training was designed to improve participants' fine motor skills of their nondominant (i.e., left) hand, with respect to writing and tracing. Daily training sessions (at home) were performed on a touch-sensitive tablet PC with a pen (12.1 in. Lenovo ThinkPad X61t, with 1024 × 768 pixels, sampling rate of 133 Hz, running under Microsoft Windows 7 Enterprise). The training program consisted of a fixed number of tasks that had to be fulfilled and were presented in random order, with 5 tasks being identical in every training session and 11 other tasks alternating to keep the training interesting. Tasks were either manually created curved paths that had to be traced as fast but also as accurately as possible, or words that had to be rewritten cursively 3 times, as fluently as possible (see Fig. 1 for examples). Both tracing of prescribed curves and writing of nonspecific words aimed at increasing the left hand's fine motor skills to control the pen and thereby presumably activating and challenging the associated cortical representational areas. Depending on each individual's speed in completing these tasks, the training lasted between 30 and 45 min per day.

A bonus system ensured continued compliance of participants. Participants received an extra amount of money of up to 360€ depending on how many daily training sessions (and MR measurements) they had actually completed. In the first 2 weeks of training, participants were scanned 3 times per week, as we assumed that the task would be most challenging in the beginning of training, which potentially could make plastic changes evolve more quickly, while in the following 5 weeks of training they were scanned twice a week. Scanning was scheduled at approximately the same time of day for every scanning point, to keep possibly occurring patterns of within-day variations of brain structure due to daytime, water consumption, or temperature as constant as possible. Further, participants were asked to not consume caffeine-containing drinks 1 h before their appointment to limit potential immediate effects of caffeine on blood flow and brain activation (Koppelstaetter et al. 2010). Once per week, participants delivered their training data (log files) on a USB drive for inspection. This was done to ensure compliance to the training regime at home.

Analysis of Behavioral Training Data

The kinematic writing and tracing data were first checked for outliers, which were excluded when exceeding 3 standard deviations, defined separately for each subject, task, and parameter. This resulted in the exclusion of 1.2% of all data points. The data were then analyzed by means of custom-written routines in Matlab R2012a (The MathWorks, Sherborn, MA, USA). The outcome variables were defined as follows:

Tracing data was quantified in terms of 1) duration (total drawing time, in seconds), 2) deviation (average distance of the drawn trace from the prescribed figure, in pixels), 3) a combined measure of speed and accuracy (arithmetic product of duration and deviation), and 4) smoothness (normalized jerk; Romero et al. 2003). Writing data was quantified in terms of 1) normalized jerk and 2) the mean number of inversions of accelerations (NIA) per stroke, where stroke segmentation was based on vertical

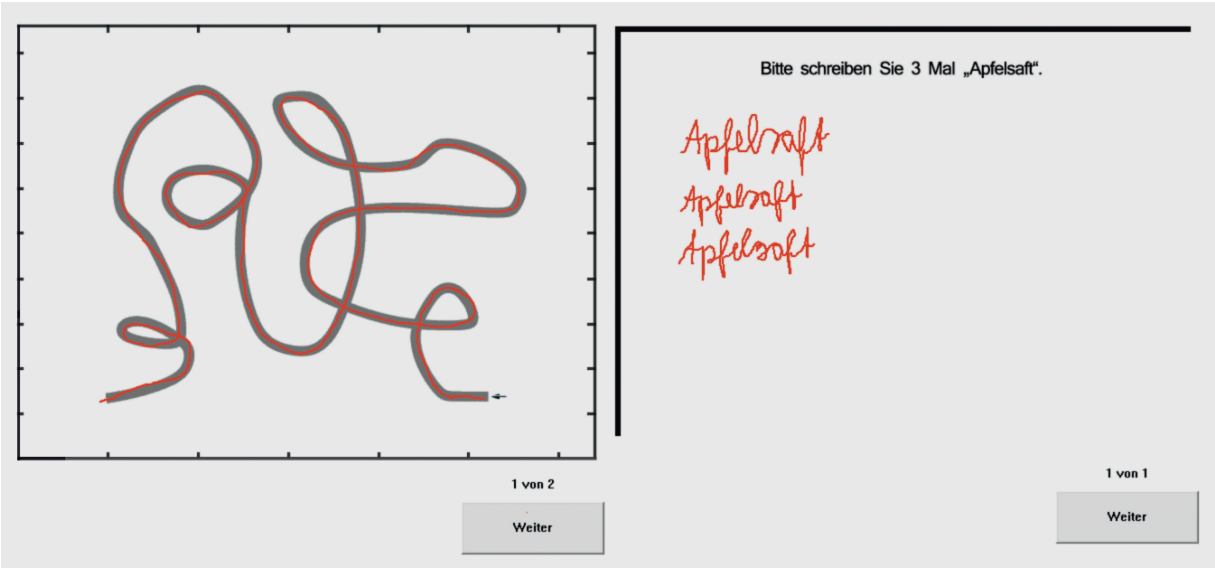

Figure 1. Behavioral training tasks. The behavioral training paradigm aimed at increasing participants' fine motor skills of the nondominant (i.e., left) hand with respect to writing and tracing. The daily training sessions (30–45 min) were performed on a touch-sensitive tablet PC with a pen and consisted of a fixed number of tasks. Tasks included curved paths that had to be traced as fast and accurately as possible, or words that had to be copied 3 times as fluently as possibly in cursive handwriting.

extrema (Marquardt et al. 1999; Hamzei et al. 2012). Both of the latter 2 measures quantify the smoothness of the writing data. For the smoothness measures, the data were first bidirectionally low-pass filtered using a finite-impulse response filter with cutoff frequency 16 Hz and order 34. We are missing writing data of 1 control participant at pretest, resulting in 30 data points for writing at pretest, compared with 31 data points for writing at posttest and 31 for tracing at pre- and posttest.

For statistical comparisons, we first normalized the behavioral data. For each task and variable, we computed a z-score across subjects, groups, and sessions of the log-transformed value and then averaged across tasks. We then ran mixed ANOVAs with the factors Time (pretest vs. posttest) and Group (experimental vs. control).

MR Image Acquisition

Structural images were collected on a Siemens Tim Trio 3T MR scanner (Erlangen, Germany) with a standard 12-channel head coil. We used a 3-dimensional T_1-weighted magnetization prepared gradient-echo sequence (MPRAGE) of 9.20 min with the following parameters: TR = 2500 ms, TE = 4.77 ms, TI = 1100 ms, flip angle = 7°, bandwidth = 140 Hz/pixel, acquisition matrix = 256 × 256 × 192, isometric voxel size = 1 mm^3. We used the prescan normalize option and a 3D distortion correction for nonlinear gradients.

Whole-brain functional images were collected using a T_2*-weighted EPI sequence sensitive to BOLD contrast (TR = 2000 ms, TE = 30 ms, FOV = 216 × 216 × 129 mm^3, flip angle = 80°, slice thickness 3.0 mm, distance factor = 20%, voxel size = 3 mm^3, 36 axial slices, using GRAPPA acceleration factor 2). Slices were acquired in an interleaved fashion, aligned to genu-splenium of the corpus callosum.

For DTI, diffusion-weighted data were collected with a 32-channel head coil on the same scanner. An EPI-based diffusion-weighted pulse sequence (full acquisition time of 25 min) was applied sequentially in 60 different directions (b = 1000 s/mm^2), with 7 nondiffusion-weighted images distributed equidistantly throughout. We applied the following parameters: TR = 11 000 ms, TE = 98 ms, FOV = 218 × 218 mm^2, 2 averages, in plane resolution 1.7 × 1.7 mm, 73 slices with a thickness of 1.7 mm using GRAPPA, acceleration factor 2, no partial k-space, and no gap.

MRI Data Processing and Analysis

Structural T_1-weighted Images

We applied voxel-based morphometry (VBM; Ashburner and Friston 2000) to the structural T_1-weighted images to quantify gray matter volume in a voxel-wise fashion. VBM is one among several available methods for quantifying structure in human MR imaging. VBM tends to be reliable (Eggert et al. 2012), and most previous studies conducted in the field of human neuroplasticity have reported VBM-based analyses (see reviews on structural plasticity, Lövdén et al. 2013; Thomas and Baker 2013). Hence, we opted for using VBM in the present study. The data preprocessing was performed using the VBM8 toolbox (Christian Gaser, University of Jena, Department of Psychiatry), implemented with SPM8 running under Matlab R2012a (The MathWorks, Sherborn, MA, USA). All structural images were visually checked for artifacts and severe motion artifacts, whereby none were detected. Using default parameters, preprocessing of the data involved bias correction, tissue classification, affine registration, DARTEL template creation, and nonlinear only modulation of gray matter segments. The resulting gray matter images were smoothed with a standard Gaussian kernel of 8 mm full-width at half maximum (FWHM). To check in how far possible results would be resistant to changes in smoothing kernel size, we also smoothed the images with a kernel of 12 mm FWHM, as large smoothing kernels reduce the rate of false positives to an acceptable level (Silver et al. 2011). This did not change our results. We refrained from any special longitudinal preprocessing of the images, such as coregistration of the images across time points, because the appropriate procedure for this has not yet been optimized for more than 2 time points (Ashburner and Ridgway 2013).

For statistical analysis of the VBM-processed images, we first restricted our analyses to only 2 time points, to simulate a typically available data set and used a flexible-factorial design with 3

factors, Subject, Time (pretest and posttest), and Group (experimental vs. control group), to test for the critical Time by Group interaction that detects differential changes for the 2 groups. Then, we applied the same model to pretest and the measurement time point in Week 4, to test again for a Time by Group interaction.

We then performed time-series analyses and used the regression framework to enter all time points of one individual into one model. We tested for linear increase of gray matter throughout the training. Alternatively, training may lead to a pronounced increase in the beginning followed by a stabilization of the structure as training continues (denoted here as inverse-quadratic-asymptotic function). Finally, we tested for a pronounced increase of structure after training onset followed by renormalization. We will refer to this shape of time course as inverse-quadratic function. In Figure 2, we illustrate the 3 potential shapes of time courses of structural alterations in the brain.

To detect regions in the brain that follow either of these 3 possible time courses, we created a separate linear regression model in SPM8 for each participant, in which we entered all VBM-preprocessed structural brain images of one participant as the dependent variable, with the respective data from the 3 different functions as a covariate. This resulted in 3 models for each participant: one model for the linear increase, one model for the inverse-quadratic-asymptotic increase, and one for an inverse-quadratic shape of change (see Fig. 2).

The linear increase model has a simple linear increase function as a covariate, based on the numbered scanning days for each individual participant (e.g., Day 1, 3, 5, 7, . . .). For the inverse-quadratic function, we used an inverse-quadratic function with the maximum at 25.5 and zeros at 1 and 50. Numbers on the x-axis correspond to days on which participants were scanned. To represent the inverse-quadratic-asymptotic shape of change, we used the same numbers as in the inverse quadratic model for the first 25 days and the constant value of the peak for the last 25 days. Importantly, for each participant, we entered those values into the model that corresponded to the exact days on which this participant was scanned, thereby ensuing an accurately specified model for each individual.

These 3 regression models were run separately for each participant, testing for the effect of the covariate. All voxels with a gray matter value below 0.2 were excluded to prevent border effects between gray and white matter. The resulting β weight images were entered in a second-level one-sample *t*-test, testing for common significant voxels for the whole group. The resulting maps were thresholded at $P < 0.001$, with a Family-Wise Error (FWE) corrected cluster extent threshold of $P < 0.05$, $k > 100$, and a correction for nonstationary smoothness. For visualization purposes, we extracted gray matter probability data from significant clusters of structural change to SPSS, using the MarsBar tool for SPM (http://marsbar.sourceforge.net/). When plotting the extracted gray matter probability data, we made use of within-subject error bars. The problem of calculating error bars in within-subject designs has received much attention in recent years (Loftus and Masson 1994; Morey 2008; Cousineau and O'Brien 2014). In within-subject designs, hypotheses generally refer to change over time and not to stable differences between participants or groups of participants. In such a situation, it is generally misleading to plot between-subject error bars. Following Cousineau and O'Brien (2014), computing error bars for within-subject designs involves 2 steps: first, centering the data to remove between-subject differences, and, second, integrating a correction factor to de-bias the standard errors obtained from the normalized data set (see Supplementary Material for the exact formulae).

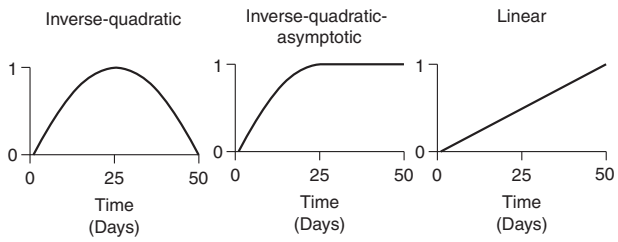

Figure 2. Three hypothetical shapes of gray matter changes. Numbers on x-axis correspond to days on which participants were scanned. The 3 functions were entered as covariates in regression models, including all VBM-preprocessed gray matter segmentations of a given participant over time as dependent variable, to test whether changes in brain regions were adequately described by any of these 3 hypothetical shapes of change in gray matter volume.

To check for a potential association between individual differences in structural brain changes and individual differences in training-related behavioral changes, we also entered individual training data as covariates into the analysis. Additionally, we inspected the correlations of individual differences in performance improvements with individual differences in gray matter changes within significant clusters 1) from pretest to Week 4; 2) from pretest to individual gray matter peaks; 3) from Week 4 to posttest; and 4) from individual gray matter peaks to posttest.

EPI Images

The fMRI data were analyzed using the SPM8 software (Wellcome Department of Cognitive Neurology, London, UK). Data processing started with slice time correction and realignment of the EPI dataset. A mean image for all EPI volumes was created, to which individual volumes were spatially realigned by means of rigid body transformations. The high-resolution structural image was co-registered to the mean image of the EPI series. Then the structural image was normalized to the Montreal Neurological Institute (MNI) template, and the normalization parameters were applied to the EPI images. A commonly applied filter of 8 mm FWHM was used. Low-frequency drifts in the time domain were removed by applying a high-pass filter cutoff of 128 s. In the single subject-level statistical analyses, we used a general linear model (GLM) to contrast right-hand writing with rest (not-writing), left-hand writing with rest, and left-hand with right-hand writing. The 6 realignment parameters were additionally entered into the model to correct for head motion. Contrast (*t*) images with the corresponding β weight images were constructed for each individual, which were then entered into the second-level analysis. Identical contrasts were formed for the fMRI finger-tapping task. Based on the second-level analysis, we constructed a whole-brain mask consisting of the activation map of the contrast left-hand writing greater than rest at pretest ($P < 0.05$, FDR-corrected), thus including all regions functionally relevant when performing the task.

We also probed for changes in the functional activation patterns over the course of training. Using the previously constructed β weights for each of the 3 contrasts (left > rest, right > rest, left > right), we built a linear regression model in which we entered all β weight images over time for one participant, with either a linear function or an inverse-quadratic function as covariate, identical to the time-series analysis of structural images. This resulted in a regression model testing for linear increase (or decrease in the inverse contrast) and a model testing for inverse-quadratic change in functional activation over time.

Additionally, we entered the β weight images for pretest and posttest in a simple paired t-test analysis ($P < 0.001$, uncorrected).

The unsmoothed preprocessed fMRI data were further analyzed with the toolbox PRoNTo (Schrouff et al. 2013), to complement the mass-univariate fMRI analyses by the use of multivariate pattern analyses (MVPA). Using machine learning-based predictive models, this pattern recognition toolbox is concerned with the automatic discovery of regularities in data through the use of computer algorithms to classify data into different categories (Schrouff et al. 2013). We focused on predicting left-hand writing versus right-hand writing with the pattern of brain activation. A Support Vector Machine (SVM) for classification was trained and tested on data from pretest, Week 4, and posttest separately, and on experimental and control groups within those clusters showing structural changes over time. We used a leave–one–subject–out cross validation scheme. Permutation testing with 100 repetitions for each classifier and mask assessed whether the classifier's accuracy was significantly different from chance level. A mixed ANOVA with the within-person factors Time (pretest vs. posttest) and Hemisphere (left vs. right) and between-person factor Group (experimental vs. control) ascertained a potential significant difference of change over time between the 2 groups.

DTI Images

The raw diffusion images were preprocessed using the FSL software package (Jenkinson et al. 2012), including corrections for possible head movement based on the nondiffusion-weighted images and inspection of image quality for all participants. The regions of interest created based on results from the VBM analyses were co-registered to the same space and then used to extract measures of MD and FA in diffusion-weighted images. Mean diffusivity values in gray matter can be used as a measure for tissue density, presumably indicating to which extent water molecules in extracellular space are able to move freely, thus providing information about tissue density (Wrigley et al. 2009).

Results

Behavioral Improvements

Mixed ANOVAs run on data from pre- and posttest assessments in the lab allowed for direct group comparisons between experimental and control groups. The analyses showed a general pattern of performance increases in the experimental group relative to the control group. There were significant time × group interactions in all tracing variables, normalized jerk, $F_{1,29} = 12.785$, $P = 0.001$, generalized eta-squared (η_G^2) = 0.063; duration, $F_{1,29} = 14.376$, $P < 0.001$, $\eta_G^2 = 0.072$; deviation by duration, $F_{1,29} = 18.387$, $P = 0.001$, $\eta_G^2 = 0.106$; except for deviation, $F_{1,29} = 0.361$, ns, $P = 0.553$, $\eta_G^2 = 0.003$. Writing variables showed the same general pattern of performance increases in the experimental group relative to the control group, for normalized jerk, $F_{1,28} = 5.172$, $P = 0.031$, $\eta_G^2 = 0.074$; a trend for statistical significance for mean number of inversions of accelerations, $F_{1,28} = 3.361$, $P = 0.077$, $\eta_G^2 = 0.031$. Importantly, all statistically reliable interactions were driven by increases in the experimental group, as shown by follow-up t-tests in the experimental group (tracing: normalized jerk $t_{(14)} = 5.978$, $P < 0.001$; duration $t_{(14)} = 6.148$, $P < 0.001$; deviation by duration, $t_{(14)} = 6.781$, $P < 0.001$; writing: mean NIA, $t_{(14)} = 3.927$, $P = 0.002$; normalized jerk writing, $t_{(14)} = 4.650$, $P < 0.001$). For illustration purposes, we display the mean behavioral learning curve of participants (see Fig. 3) as well as an example of single-subject data (Fig. 4).

There were no changes in right-hand performance selectively for the experimental group in writing parameters, that is, no significant time × group effects in right-hand writing parameters ($Fs_{1,28} < 0.016$, ns). However, we observed some significant time × group effects in right-hand tracing parameters, namely for normalized jerk, $F_{1,29} = 4.824$, $P = 0.036$, $\eta_G^2 = 0.025$, duration, $F_{1,29} = 5.984$, $P = 0.021$, $\eta_G^2 = 0.033$, and deviation, $F_{1,29} = 5.638$, $P = 0.024$, $\eta_G^2 = 0.039$. This is consistent with previous reports of transfer from a skill trained with the nondominant to the dominant hand (Grafton et al. 2002).

Gray Matter Time by Group Interaction Effects

Analyses of the VBM-processed T_1-weighted MR images revealed differential increases in gray matter structure for the experimental group compared with the control group in the form of a time by group interaction in left and right primary motor cortices after 4 weeks of training (i.e., pretest to Week 4; peaks at −16 −27 60 and 24 −25 69). The effect maps, using a common threshold of $P < 0.001$, with an FWE corrected cluster extent threshold of $P < 0.05$ ($k > 100$), including correction for nonstationary smoothness, are depicted in Figure 5A. In contrast, after 7 weeks of training (i.e., pretest to posttest), we found no significant interaction, even at a more lenient threshold of $P < 0.01$ (uncorrected) (see Supplementary Fig. 1 for change scores from pretest to Week 4 and from pretest to posttest for individual participants of both groups).

Time Course of Gray Matter Changes: Time-Series Analyses

We then analyzed all time points acquired for each participant in the experimental group and tested for the 3 different shapes of gray matter changes during training. For the linear-increase function, the regression results showed significant effects, using the same stricter threshold as above, in left putamen, left temporal lobe, and left and right cerebellum. Fitting an inverse-quadratic-asymptotic function to the data also revealed effects in left putamen, left inferior temporal lobe, and left cerebellum, and additionally showed effects in right putamen and left primary motor cortex. The inverse-quadratic function again yielded an effect in left inferior temporal lobe, and additionally a cluster in right primary motor cortex. That is, right primary motor cortex showed an increase of gray matter volume followed by renormalization. A trend for this increase-normalization pattern was also found in left primary motor cortex and right putamen when thresholding at a P value of 0.001 (uncorrected, without FWE-corrected cluster extent threshold).

Analyses of functional MR images (fMRI) acquired during the writing task showed higher levels of BOLD signal during writing than during rest in the regions that showed significant structural changes in left and right primary motor cortices and in putamen, suggesting that these regions were indeed involved in writing (left-hand writing vs. rest, FDR-corrected $P < 0.05$; Fig. 5B i and ii). Moreover, the regions showing significant structural change in left (−38 −13 46, BA 4) and right (40 −18 51, BA 4) primary motor cortices overlapped with the functional activation maps during finger tapping and were located in close proximity of the anatomical "hand knobs" (Yousry et al. 1997) (see Fig. 6). Given this involvement, and given the greater motor cortex increase in the experimental group than in the control group, we conducted follow-up analyses targeting the regions in left and right motor cortices and putamen that showed reliably greater activity during writing than during rest (see Supplementary Table 1 for a summary of spatial coordinates, model type, and model fit R^2).

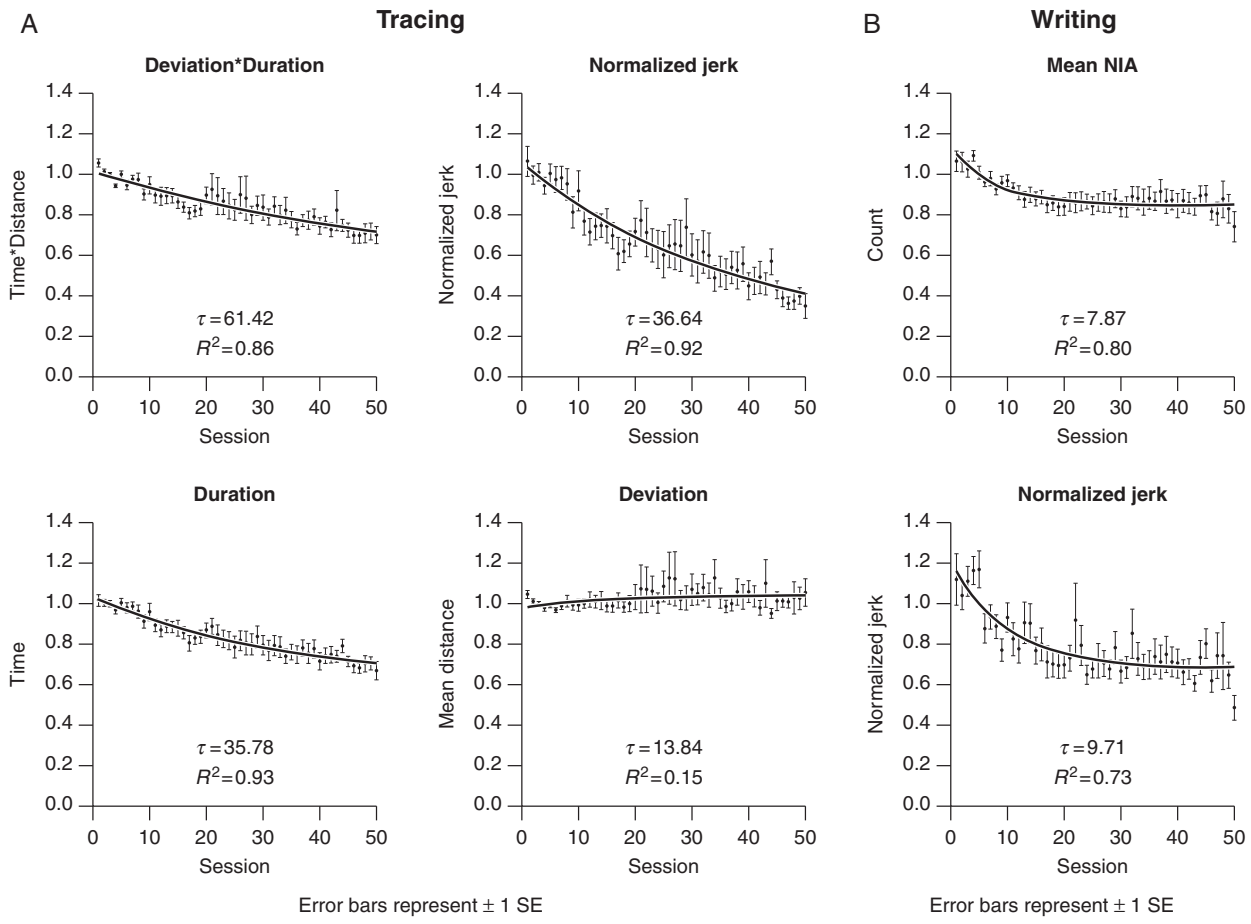

Figure 3. Improvements in left-hand performance. On average, participants' performance improved significantly on tracing tasks (on the left side) and on writing tasks (on the right side). Tracing data (*A*) was quantified in terms of a combined measure of speed and accuracy (arithmetic product of duration and deviation), smoothness of the tracing movement (normalized jerk), duration (total drawing time, in seconds), and deviation (average distance of the drawn trace from the prescribed figure, in pixels). Writing data (*B*) was quantified in terms of the mean number of inversions of accelerations (NIA) per stroke, where stroke segmentation was based on vertical extrema, and normalized jerk. Fitting of exponential curves to individual training data assessed training effects for each of these variables. Fitted training data were normalized with respect to the first 5 training session. Hence, initial values for all variables are around 1, and decreases from this initial value correspond to performance improvements. Training effects were quantified by the time constant τ of the exponential fit, indicating how fast participants approached the estimated asymptote, as well as the relative improvement R^2, which is the fitted value at the last session. Data shown here are averaged across all participants and are displayed with error bars representing 1 standard error.

For visualization purposes, we extracted gray matter probability values of significant clusters in left and right primary motor cortices and plotted those values averaged per training week (Fig. 7; see Supplementary Fig. 2 for all data points). We used those significant clusters resulting from the fitting of the inverse-quadratic-asymptotic function in left motor cortex and the inverse-quadratic function for right motor cortex. As Figure 7 shows, there was a 2.20% volume increase (Week 4) in right primary motor cortex, a region in which changes were captured only by the inverse-quadratic function. In line with the fitted function, this effect almost renormalized to baseline level towards the end of training (+0.64%). The region in left motor cortex, detected with the inverse-quadratic-asymptotic function, exhibited a later peak in Week 7, with a 2.72% volume increase, which descended to 1.29% at posttest (see Fig. 7). This later decrease may explain why, using a more lenient significance threshold, a comparable region in left precentral gyrus was also detected with the inverse-quadratic function.

DTI showed no significant change in mean diffusivity values (right motor cortex: linear: $P = 0.796$, quadratic: $P = 0.349$; left motor cortex: linear: $P = 0.685$, quadratic: $P = 0.740$), suggesting that the structural gray matter alterations did not entail significant changes in tissue density (Wrigley et al. 2009; Lövdén et al. 2013). In left motor cortex, there was a decrease in fractional anisotropy over time (linear: $F_{1,8} = 8.407$, $P = 0.020$; quadratic: $F_{2,7} = 4.048$, $P = 0.068$). In right motor cortex, there was only a trend towards a decrease in fractional anisotropy over time (linear: $F_{1,8} = 4.766$, $P = 0.061$; quadratic: $F_{2,7} = 2.846$, $P = 0.125$). Up to this point, it remains unclear what fractional anisotropy values in gray matter represent.

The detected structural changes in left and right putamen (peaks at –34 14 –2; 24 11 –2) were within the areas activated in the writing task as well. Therefore, we also extracted gray matter probability values for these regions using significant clusters from the inverse-quadratic-asymptotic function (Fig. 7*C*,*D*). Right putamen showed a 6.20% increase to Week 4 of training. Consistent with the extracted values, gray matter structure in right putamen followed the inverse-quadratic function on a more lenient significance threshold, indicating a trend towards partial renormalization for this region as well, with gray matter probability decreasing to 4.02% at posttest. Similar to the effect in left primary motor cortex, the effect in left putamen, detected

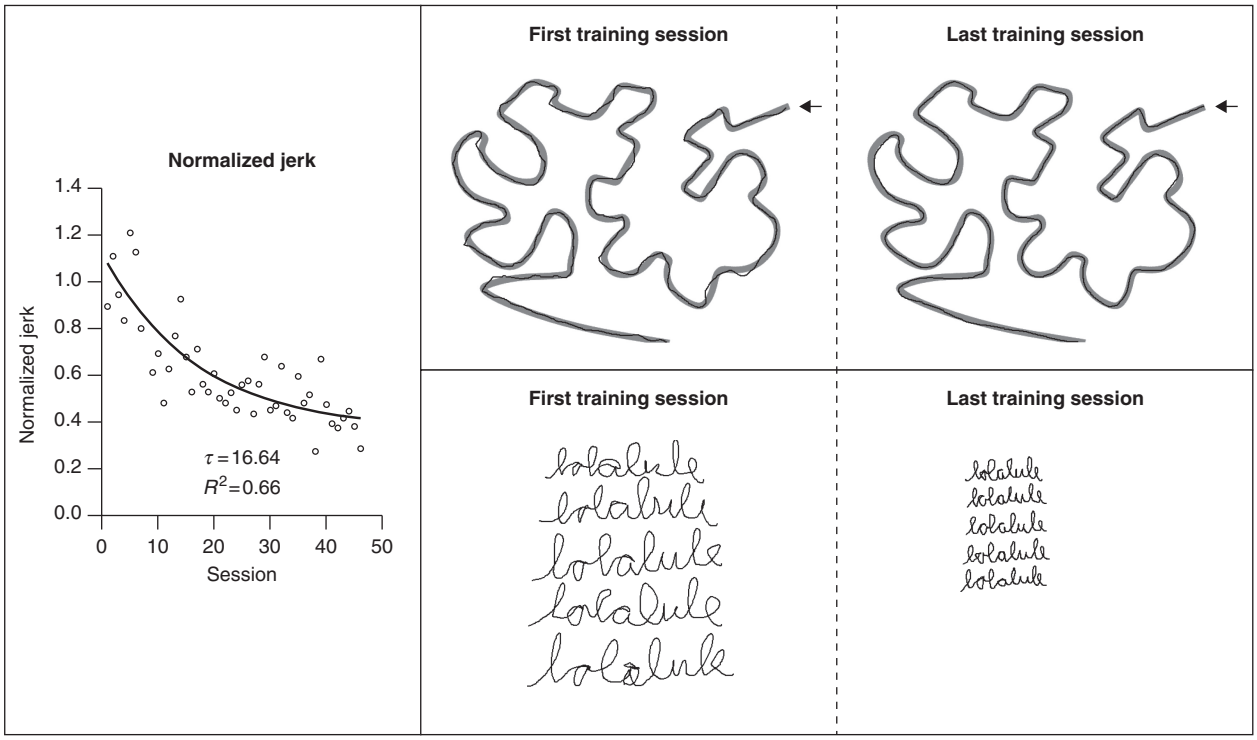

Figure 4. Examples of behavioral data. For 1 subject, normalized jerk data as a measure for smoothness of tracing, throughout the training over 7 weeks is displayed (left panel), and, from the same subject, examples of tracing and writing at the first and last training session.

both by the linear increase and inverse-quadratic-asymptotic function, peaked late in Week 7 (3.91%) and descended again to posttest (2.15%).

Correlations Between Structural Changes and Behavioral Changes

We did not observe any statistically reliable correlations between individual differences in structural brain changes and individual differences in training-related performance improvement, neither at the whole-brain level, nor in relation to the extracted gray matter probabilities from within significant regions identified by time-series analyses.

Functional Reorganization Within Left and Right Motor Cortex

We did not find significant large-scale functional activation changes in overall writing-related functional activity over time ($P < 0.001$, uncorrected; neither in time-series analysis nor as analyzed with paired t-tests from pretest to Week 4 or posttest). These results are consistent with the assumption that learning takes place in the absence of macroscopic functional reorganization in regions that are more activated during left-hand writing than rest. This view does not exclude more fine-grained changes within these regions. To check for functional redistribution of activation within the regions showing significant structural change over time, we performed multi-voxel pattern analyses (Schrouff et al. 2013) on voxels in regions of structural change in left and right primary motor cortices as revealed by our time-series analyses. We classified left-hand writing against right-hand writing at pretest, in Week 4, and at posttest in separate analyses.

At pretest, the classifier performed above chance level for subjects in the experimental group, in both left (balanced accuracy: 56.3%, $P = 0.01$) and right motor cortex (60.0%, $P = 0.01$, see Fig. 8). In Week 4, when gray matter volume had increased considerably, classification rates were still significantly above chance level for right-hand writing (in both left and right motor cortex at 56.2%). At posttest, however, the classifier remained at chance level (left motor cortex: 51.0%, $P = 0.42$; right motor cortex: 48.7%, $P = 0.76$). This decrease in accuracy results from pretest to posttest (right motor cortex: $t_{(14)} = 2.921$, $P = 0.011$; left motor cortex: $t_{(14)} = 1.617$, $P = 0.128$, *ns*) suggests that the pattern of activation in right motor cortex and to some extent also in left motor cortex, as detectable by our methods, contained less information encoding differences between left-hand and right-hand writing after training. Even when both significant clusters in left and right motor cortex were combined and used simultaneously to classify left-hand writing from right-hand writing, the same pattern emerged: left-hand writing was classifiable from right-hand writing, at pretest (54.8%, $P = 0.05$) and in Week 4 of training (57.8%, $P = 0.01$), but no longer at posttest (49.3%, ns). This change in classification accuracies was selective for the experimental group, as shown by a significant time × group interaction, $F_{1,29} = 5.208$, $P = 0.030$, $\eta^2 = 0.0009$, both in left and in right motor cortex. The 3-way interaction (i.e., time × group × hemisphere) was not significant, $F_{1,29} = 0.149$, $P = 0.702$, ns. At posttest, classification accuracy for the training group did not differ from chance at posttest (left motor cortex: 51.0%, $P = 0.42$; right motor cortex: 48.7%, $P = 0.76$). Apparently, extensive left-hand training rendered the activation pattern during left-hand writing more similar to right-hand writing, eliminating classifiable differences between the 2 activation patterns. The classifier was also administered using the masks from the significant clusters in left and right putamen, respectively. The classification did not exceed chance level for these regions, neither at pretest nor in Week 4 or at posttest. Classification accuracies for the control group were continuously above chance level in left motor cortex

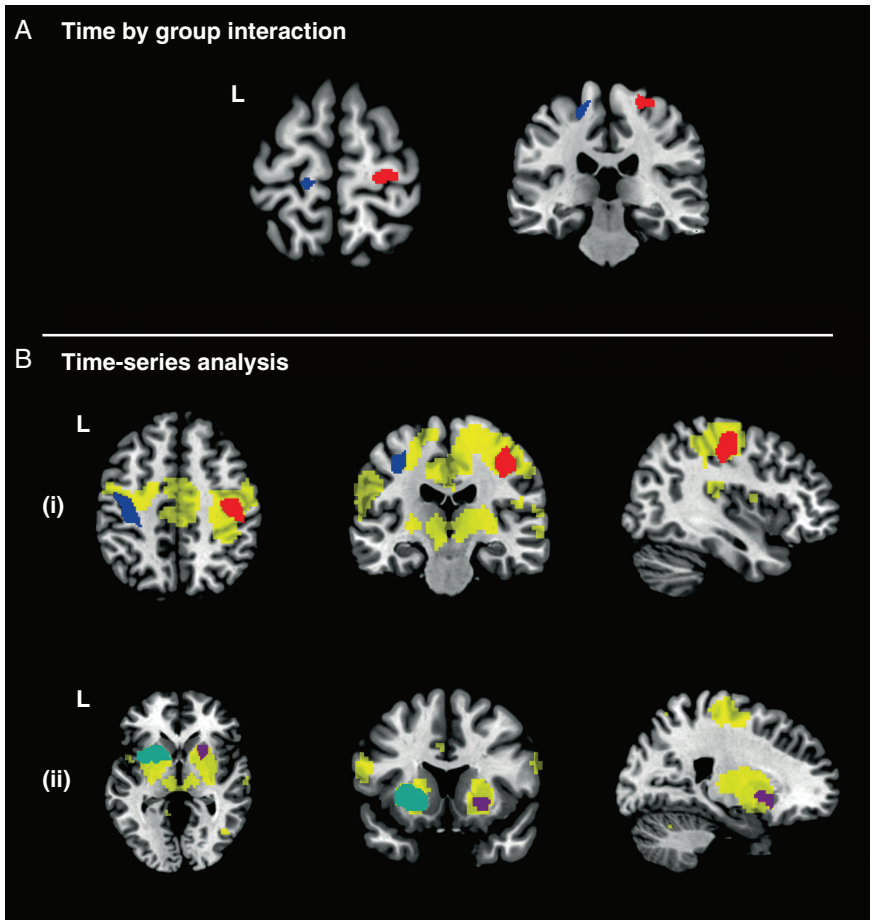

Figure 5. Gray matter changes accompanying motor skill acquisition. (*A*) When comparing pretest with Week 4, the Time by Group interaction was statistically reliable, indicating that increases in gray matter structure were greater in the experimental group than in the control group in both left (blue; peak at −16 −27 60; cluster comprises 173 voxels) and right (red; 24 −25 69; cluster comprises 160 voxels) primary motor cortices. When comparing pretest with posttest (i.e., after 7 weeks of training), no reliable Time × Group interactions were observed for any brain region. (*B*) Using time-series analyses, we found regions of significant volume change in (i) left and right primary motor cortices (depicted in red and blue; −38 −13 46; 40 −18 51) and (ii) left and right putamen (in light blue and violet; −34 14 −2; 24 11 −2). Functional activation during left-hand writing is depicted in yellow. Results are shown for the fitted inverse-quadratic-asymptotic function in left motor cortex and bilateral putamen and for the fitted inverse-quadratic function in right motor cortex.

(57.5, 53.0, and 58.6% for pretest, Week 4, and posttest, respectively). However, in right motor cortex, classification accuracies of the control group were only above chance level at pretest (55.3%) and failed to be significant in Week 4 and at posttest (52.7 and 53%), even though there was no significant change in classification accuracies from pretest to posttest, as assessed with paired *t*-tests on the individual accuracies of each subject (left motor cortex: $t_{(15)} = 0.83$, $P = 0.42$; right motor cortex: $t_{(15)} = -0.33$, $P = 0.75$). We report classification accuracies for both experimental and control groups with 95% bootstrapped confidence intervals in Supplementary Table 2.

Discussion

We report that individuals training fine motor skills of writing and tracing with their nondominant left hand displayed significant expansion of gray matter volume of both left and right primary motor cortex relative to a control group. These group-differential increases in volume were significant after 4 weeks, but not after 7 weeks of practicing. In line with this result, time-series analyses of the structural MR images of participants in the experimental group revealed that changes of gray matter in right primary motor cortex followed an inverse-quadratic function. That is, an initial expansion was followed by partial renormalization, despite continued practice and increasing task proficiency.

Analogous patterns of change have been observed in animals, in related manifestations of plasticity. For example, new dendritic spines formed rapidly in mice training a reaching task (Xu et al. 2009). This rapid increase was followed by a slower process of elimination of spines that had existed before training, returning the overall number of spines to almost pretraining levels, while performance remained high. Similarly, monkeys and rats learning to retrieve food exhibited training-related gray matter volume or cortical map expansion that partially renormalized while behavioral performance remained stable (Molina-Luna et al. 2008; Quallo et al. 2009). Effects of exercise on progenitor cell proliferation have also been shown to follow an inverted u-shape (Kronenberg et al. 2006; Overall et al. 2013). On a different time scale, the pruning model of brain maturation is another example of the same pattern of plastic changes: increase in the number of synapses followed by experience-dependent selective stabilization of important connections and the elimination of others (Edelman 1987; Changeux and Dehaene 1989). Thus, expansion followed by partial renormalization may be a common principle that unites different manifestations of plasticity.

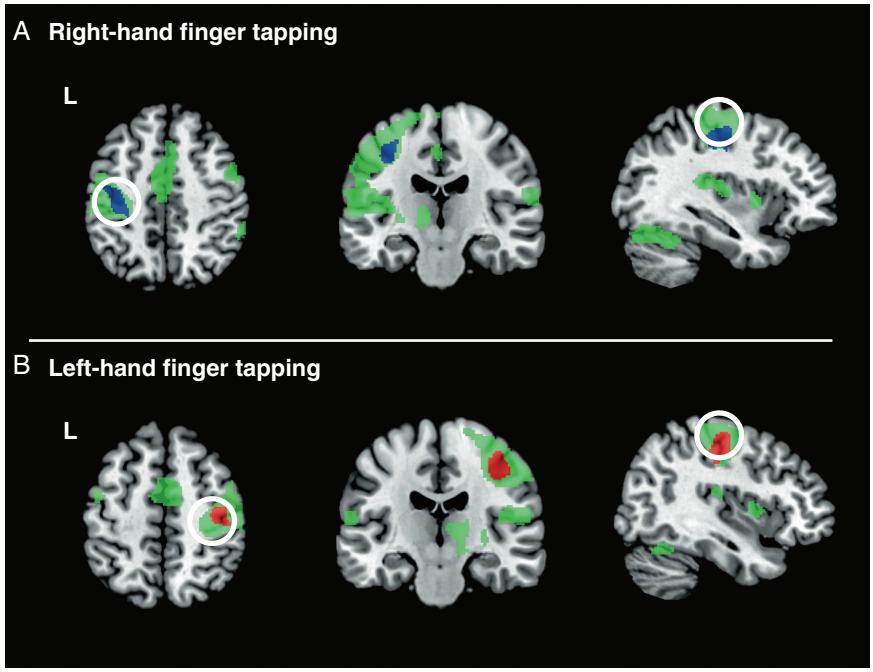

Figure 6. Functional activation maps for finger tapping and anatomical hand knobs. There is high spatial congruence between anatomical hand knobs (marked with white circles) and structural change (displayed in blue and red for left and right motor cortices). The functional activation maps during (*A*) right-hand finger tapping and (*B*) left-hand finger tapping are depicted in green.

The observed growth of primary motor cortices is consistent with the important role of these regions in the acquisition of fine motor skills (Sanes and Donoghue 2000; Ungerleider et al. 2002). The fact that structural changes were found in both hemispheres is in line with our functional imaging results during left-hand writing, which showed activation in both left and right primary motor cortex, and with the established relevance of both the contralateral and ipsilateral hemisphere for performing motor tasks with the nondominant hand (Kim et al. 1993; Li et al. 1996). We note, however, that growth of left primary motor cortex did only show a trend for renormalization using a more lenient significance threshold during the measured period of training. It is possible that different regions follow the same overall shape of practice-related change but at different paces, so that we might have seen a more pronounced partial renormalization phase in left motor cortex as well if the training and scanning regime had been continued. A closer look at the extracted time course in left motor cortex shows that volume changes are actually in line with this trend for partial renormalization. Absence of training-related changes after 7 weeks of training in our pretest–posttest comparison also points in the direction of renormalizing structure as training continues.

In general, an expansion-renormalization process is presumably a more efficient way for the brain to reorganize and adjust than a constant growth process. Otherwise people would be unable to learn new skills while maintaining earlier skills, as there would be constant competition between adjacent brain regions. Future studies need to follow up on this finding to eventually resolve whether practice-related expansion followed by partial renormalization is a general phenomenon across regions. More studies on structural plasticity with at least 3 measurement time points are needed to investigate if this pattern also occurs following training of other tasks than writing and drawing with the nondominant hand.

Among the regions functionally active during our fMRI left-hand writing task, left and right putamen also showed significant growth according to time-series analysis in practicing individuals. This finding is expected given that the basal ganglia have been repeatedly associated with motor learning (Hikosaka et al. 2002; Doyon et al. 2009). The finding of growth in putamen should however be considered preliminary, because—in contrast to clusters in left and right motor cortex—changes in putamen did not reliably differ from changes in the control group in the whole-brain interaction analysis. For the same reason, and given that they did not show signs of functional involvement during writing, the structural changes observed in cerebellum and left temporal lobe of the experimental group also need to be interpreted with caution. Nevertheless, it is worth noting that cerebellar movement-related functions are solidly established (Hikosaka et al. 2002). The cerebellum plays an important role in motor control, is known to contribute to coordination, precision, and accurate timing, and is necessary for several types of motor learning (Ungerleider et al. 2002; Penhune and Doyon 2005; Mottolese et al. 2013). In several other—though cross-sectional—studies, associations between structural alterations in cerebellar gray matter volume and various kinds of motor skills have been observed (Gaser and Schlaug 2003; Bermudez and Zatorre 2005; Cannonieri et al. 2007; Han et al. 2009; Park et al. 2009).

In the present study, we also observed training-induced changes in left temporal lobe (fusiform gyrus, inferior/superior temporal gyrus, parahippocampal gyrus). These changes may be related to language processing (Binder et al. 1997), as our left-hand writing task required participants to write actual words with their left hand, something that they were used to do automatically for years with their right, dominant hand. Alternatively, it has been argued that regions in temporal cortex are associated with memory processes that provide a representation of the learnt skill that is less closely tied to an action system,

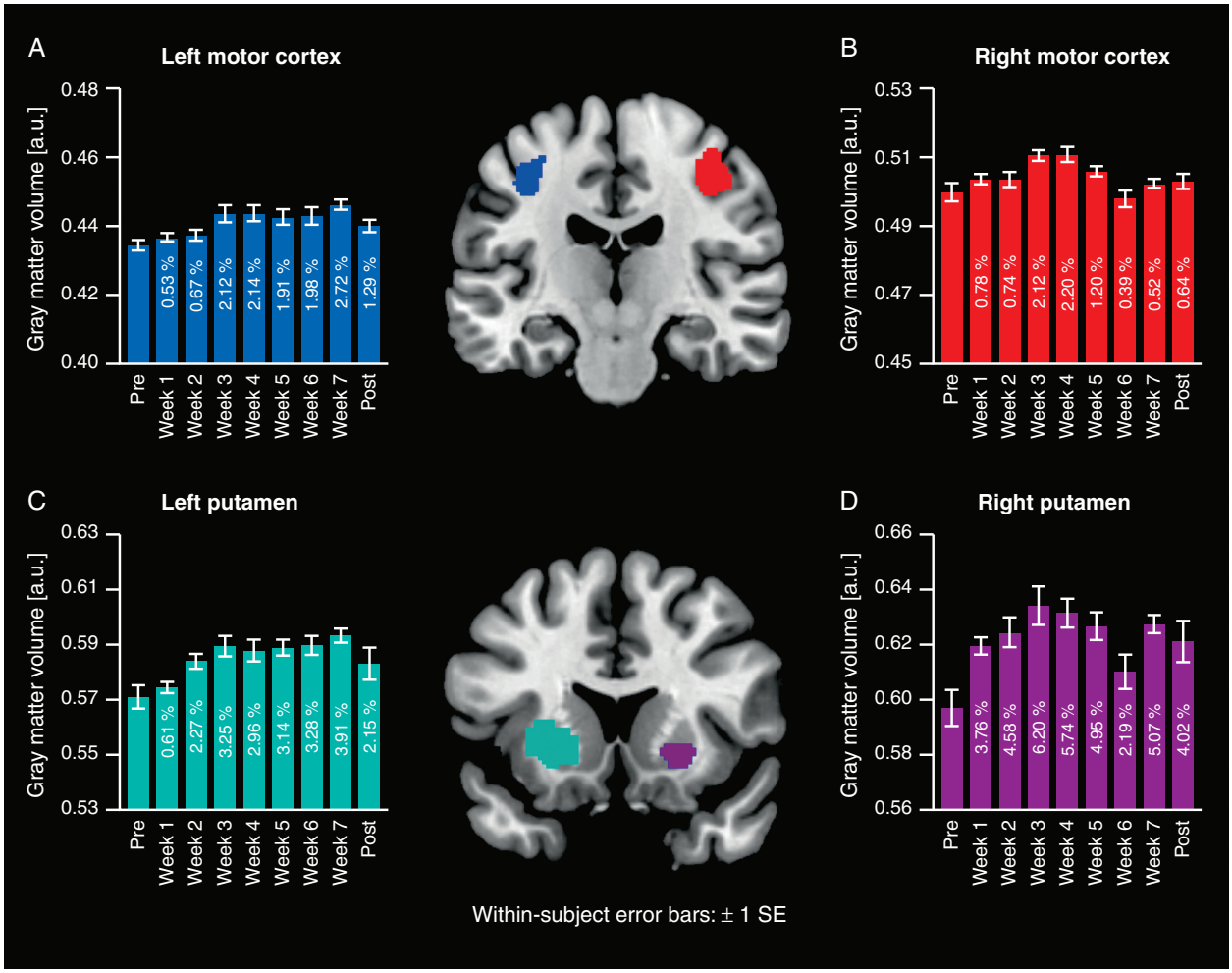

Figure 7. Extracted gray matter volumes as a function of training week. Volumes represent weekly averages. Error bars represent standard error (SE) at each time point after removing between-person variability (Cousineau and O'Brien 2014).

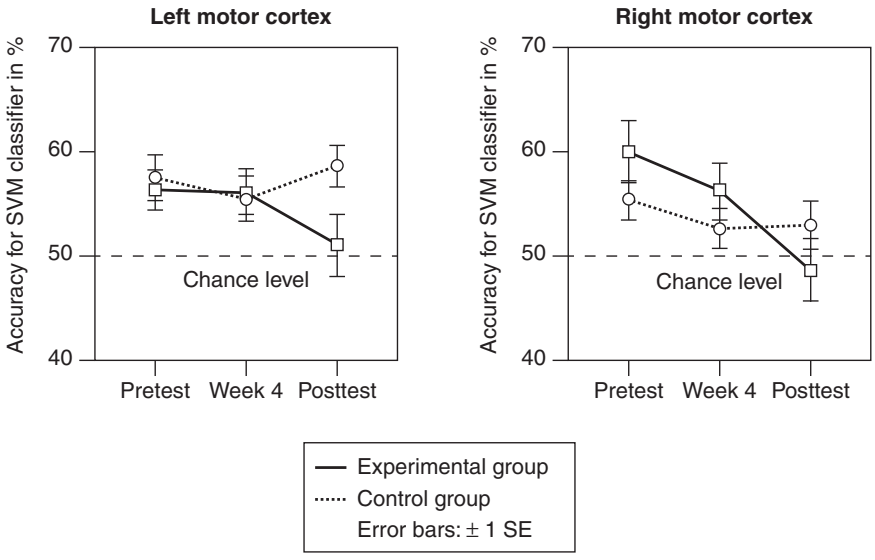

Figure 8. Functional reorganization within motor cortex. Balanced accuracies of a support-vector machine (SVM) classifying left-hand writing versus right-hand writing within the regions of structural change. After 7 weeks of training, left-hand writing was no longer reliably separable from right-hand writing in either left or right motor cortices. Classifier accuracies for the control group did not vary across sessions. Error bars represent 1 standard error.

especially when the learning is explicit (Grafton et al. 2002). In line with this hypothesis, inferior temporal regions have been found to show a learning-related increase in regional cerebral blood flow (Grafton et al. 1995; Hazeltine et al. 1997). Future research is needed to test and disambiguate these interpretations.

The observed structural changes in the primary motor cortices are particularly interesting in light of the functional reorganization of these regions as revealed by MVPA. Specifically, the MVPA results indicate that patterns of activation in left and right motor cortex, as discernible by the present resolution of MR images, discriminate less well between left-hand and right-hand writing after training than before training. Most of the functional remodeling underlying this loss in discriminability presumably took place between Week 4 and posttest, which is the period during which declines in volume were observed. The observed coincidence between functional reorganization and structural renormalization is consistent with the view that selection plays a key role during later stages of experience-dependent cortical change.

In contrast to these functional changes within the regions of primary motor cortices displaying structural change, we observed no overall, large-scale changes of functional activity elicited by left-hand training. This lack of large-scale changes theoretically speaks against the possibility that the time course of structural gray matter changes is influenced by large-scale functional reorganization. Obviously, it is impossible to exclude this option entirely, though, since null effects are difficult to interpret. Additionally, the MR scanner provides an environment to perform a writing task that differs markedly from the regular writing environment outside the scanner. Even though activation of the core regions involved in the execution of the task will be detected (Harrington et al. 2007), it remains unclear whether changes in the overall network that are due to a more skilled or automatized execution of the task as trained outside the scanner can even be picked up by this task inside the scanner. Future analyses of both univariate and multivariate nature focusing on changes in the overall functional connectivity within the motor network including motor cortices, basal ganglia, and cerebellum may yield results that expand and qualify the structural and functional findings reported in this manuscript.

We did not observe any statistically reliable correlations between individual differences in structural brain changes and individual differences in behavioral changes. Recent reviews on structural plasticity have emphasized that structural plastic change ought to be relevant for behavior to be considered meaningful (Thomas & Baker 2013). Such evidence for behavioral relevance is often equated with correlations between structural increases and performance increases based on between-subject variation within the training group. In our view, this focus on between-person variation may not always be warranted. It assumes that a given amount of increase in gray matter has comparable behavioral "consequences" across participants, despite the fact that participants start the intervention at different "levels" of gray matter volume and behavioral performance. This assumption of a linear interval metric that applies to all participants may not correspond to reality, exactly because individuals differ widely in their cerebral and behavioral starting points.

Structural changes in gray matter as discernible with MRI and quantifiable with VBM are generally not the result of a single process operating in isolation. Instead, such changes are likely to result from a more or less coordinated number of alterations that involve different cell types and represent a conglomerate of synaptogenesis, changes in neuronal morphology, axon sprouting, dendritic branching, glial changes, and angiogenesis (for a summary, see also Zatorre, Fields, and Johansen-Berg 2012). The time course of glial growth and retraction in response to behavioral manipulations is not well understood and may obfuscate the neural contributions to changes in regional volume (Anderson et al. 1994), thereby further complicating associations between gray matter changes and behavioral changes.

The nonlinear time course of gray matter increases accompanying motor skill acquisition, as identified in the present study, complicates rather than simplifies the search for plasticity-related brain-behavior associations further. Given the curvilinear shape of change, it is likely that the specific time points chosen for analysis influence the association between gray matter changes and behavioral changes. Differences in learning may be related to individual differences in the rate of initial gray matter increase, in the rate of later decrease, or with a combination of both. In addition, correlations between behavior and structure may well be time-lagged, such that structural changes precede changes in behavior, or vice versa. More specific hypotheses and more fine-grained physiological and behavioral methods are needed to examine these options in greater details (Bernal-Rusiel et al. 2012, 2013).

A limitation of this study is the passive control group, which received fewer scans than the experimental group. It is conceivable that one could have scanned the control group exactly as often as the experimental group, and that one could have asked them to perform some random activity on a tablet computer for a matched amount of time, thereby turning them into an active control group rather than a passive one. While the regional specificity of results makes us confident to rule out expectancy effects or social interaction effects in the experimental group, future studies may investigate "normal" daily fluctuations in MR images over a comparable amount of time to better characterize variability and reliability of common MR methods.

Our decision to train participants for a period of 7 weeks was based on previous studies in the domain of motor skill acquisition, including our own behavioral pilot data, which suggested that participants' learning curve would converge towards an asymptote after about 4–5 weeks of training. As we were interested in delineating structural changes during skill acquisition up to asymptotic performance, we opted for a 7-week training period. In retrospect, it would have been interesting to continue training beyond this time period to find out whether left motor cortex, similar to right motor cortex, would also show a pattern of increase followed by renormalization, albeit in a more protracted manner.

Another limitation of the reported study concerns the confounding factors that influence a volume measure derived from MRI, in general, and from VBM, in particular. It is possible that the positioning of participants within the head coil and scanner influences gray matter volume as captured by MR images, influencing how the automatic software packages label them. As positioning inside the head coil influences the intensity homogeneity of the images, it can have a negative impact on the segmentation results if the head was not positioned in the very center of the coil. 3D distortion correction for nonlinear gradients is crucial for reliable volume measures, and volume measurements are believed to be more accurate if the region-of-interest is not farther than 10 cm from the isocenter of the magnet. Another confounding factor might be water or caffeine consumption of participants before image acquisition (Duning et al. 2005; Vidyasagar et al. 2013). Therefore, in this study, participants' consumption of caffeine-containing drinks was assessed on the day of scanning, and participants were asked to not consume any caffeine within the last hour before scanning to avoid

any possible immediate effects on measurement. In any case it appears unlikely that changed blood flow would have contributed or confounded the results reported here in any major way, as the effects we found were region specific (close to the anatomical hand knob and in the basal ganglia). It is hard to imagine reasons why changes in blood flow would be restricted to these motor-related areas and would show a consistent temporal pattern over 18 scanning sessions. At a more general level, hormonal influences on gray matter structure are another potential confound. It has been shown previously that estrogen levels are associated with hippocampal and parahippocampal volumes (Pletzer et al. 2010; Lisofsky et al. 2015). To exclude hormonal effects related to women's normal menstrual cycle, only men were included in this study.

Training outcomes and effects on the brain might be highly dependent on the motivation of a given participant, the personal relevance of the to-be-learned skill, and the effort invested in training. Small sample size is a further limitation, though the high number of scans per person yielded a more reliable and valid picture of training-related changes in brain structures than previous studies in this field.

An additional point to consider is the training length and intensity, as well as the initial proficiency level of participants in the beginning of training. It remains to be investigated which time course gray matter alterations follow when training is continued beyond 7 weeks of training. A replication and closer look at "turning points" in structural change, that is, points at which initial increases start to renormalize, and their associations to learning rate, attained proficiency, and invested effort will provide further insight into the etiology and temporal dynamics of gray matter changes.

In summary, we investigated the time course of changes in human gray matter volume in response to motor training, as measured with MRI, in much greater detail than ever before. The data support a nonlinear view on human neuroplasticity, suggesting that an initial expansion of structure may be followed by partial renormalization, despite continued practice and performance gains. These findings may explain why previous studies with shorter practice periods sometimes have revealed larger structural changes than longer and more extensive training regimes. If we had only collected measurements before and after the 7-week practice period, we would have failed to detect any structural MR changes. Our results are crucial for future studies on other unknown aspects of experience-dependent structural changes observed with MR imaging in humans, such as the biological mechanisms behind changes in the MR signal (Zatorre et al. 2012) and their behavioral correlates (Lövdén et al. 2013; Sampaio-Baptista et al. 2014). Only more complex study designs with at least 3 or ideally even more measurement time points over the course of training may be able to appropriately capture the process of training-induced plastic changes in brain structure. Future studies should investigate the generality of the expansion-partial-renormalization-pattern across functional domains, training duration, brain regions, and age groups.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This study was funded by the Max Planck Institute for Human Development and the Gottfried Wilhelm Leibniz award 2010 of the German Research Foundation given to U.L. During the work on her dissertation, E.W. has been funded by the International Max Planck Research School on the Life Course (LIFE). M.L. was supported by a "distinguished younger researcher" grant from the Swedish Research Council (446-2013-7189).

Notes

The authors thank all participants and student assistants. Thanks also to Claudio Brozzoli for comments on an earlier draft of this paper. The authors declare no competing financial interests. *Conflict of Interest*: None declared.

References

Anderson BJ, Li X, Alcantara AA, Isaacs KR, Black JE, Greenough WT. 1994. Glial hypertrophy is associated with synaptogenesis following motor skill learning, but not with angiogenesis following exercise. Glia. 11:73–80.

Ashburner J, Friston KJ. 2000. Voxel-based morphometry—The methods. Neuroimage. 11:805–821.

Ashburner J, Ridgway GR. 2013. Symmetric diffeomorphic modeling of longitudinal structural MRI. Front Neurosci. 197:1–19.

Bellander M, Berggren R, Mårtensson J, Brehmer Y, Wenger E, Li T-Q, Bodammer NC, Shing Y-L, Werkle-Bergner M, Lövdén M. 2016. Behavioral correlates of changes in hippocampal gray matter structure during acquisition of foreign vocabulary. Neuroimage. 131:205–213.

Bermudez P, Zatorre RJ. 2005. Differences in gray matter between musicians and nonmusicians. Ann N Y Acad Sci. 1060:395–399.

Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR. 2012. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. Neuroimage. 66C:249–260.

Bernal-Rusiel JL, Reuter M, Greve DN, Fischl B, Sabuncu MR. 2013. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. Neuroimage. 81:358–370.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. 1997. Human brain language areas identified by functional magnetic resonance imaging. J Neurosci. 17:353–362.

Cannonieri GC, Bonilha L, Fernandes PT, Cendes F, Li LM. 2007. Practice and perfect: Length of training and structural brain changes in experienced typists. Neuroreport. 18:1063–1066.

Changeux JP, Dehaene S. 1989. Neuronal models of cognitive functions. Cognition. 33:63–109.

Cousineau D, O'Brien F. 2014. Error bars in within-subject designs: a comment on Baguley (2012). Behav Res Methods. 46:1149–1151.

Doyon J, Bellec P, Amsel R, Penhune V, Monchi O, Carrier J, Lehéricy S, Benali H. 2009. Contributions of the basal ganglia and functionally related brain structures to motor learning. Behav Brain Res. 199:61–75.

Doyon J, Benali H. 2005. Reorganization and plasticity in the adult brain during learning of motor skills. Curr Opin Neurobiol. 15:161–167.

Draganski B, Gaser C, Busch V, Schuierer G, Bogdahn U, May A. 2004. Changes in grey matter induced by training. Nature. 427:311–312.

Draganski B, Gaser C, Kempermann G, Kuhn HG, Büchel C, May A. 2006. Temporal and spatial dynamics of brain structure changes during extensive learning. J Neurosci. 26:6314–6317.

Driemeyer J, Boyke J, Gaser C, Büchel C, May A. 2008. Changes in gray matter induced by learning—Revisited. PLoS One. 3:1–5.

Duning T, Kloska S, Steinsträter O, Kugel H, Heindel W, Knecht S. 2005. Dehydration confounds the assessment of brain atrophy. Neurology. 64:548–550.

Edelman GM. 1987. Neural darwinism: the theory of neuronal group selection. New York (NY): Basic Books.

Eggert LD, Sommer J, Jansen A, Kircher T, Konrad C. 2012. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. PLoS One. 7(9):e45081.

Fu M, Zuo Y. 2011. Experience-dependent structural plasticity in the cortex. Trends Neurosci. 34:177–187.

Gaser C, Schlaug G. 2003. Brain structures differ between musicians and non-musicians. J Neurosci. 23:9240–9245.

Gelnar PA, Krauss BR, Sheehe PR, Szeverenyi NM, Apkarian AV. 1999. A comparative fMRI study of cortical representations for thermal painful, vibrotactile, and motor performance tasks. Neuroimage. 10:460–482.

Grafton ST, Hazeltine E, Ivry R. 1995. Functional mapping of sequence learning in normal humans. J Cogn Neurosci. 7: 497–510.

Grafton ST, Hazeltine E, Ivry RB. 2002. Motor sequence learning with the nondominant left hand. Exp Brain Res. 146:369–378.

Granert O, Peller M, Gaser C, Groppa S, Hallett M, Knutzen A, Deuschl G, Zeuner KE, Siebner HR. 2011. Manual activity shapes structure and function in contralateral human motor hand area. Neuroimage. 54:32–41.

Gryga M, Taubert M, Dukart J, Vollmann H, Conde V, Sehm B, Villringer A, Ragert P. 2012. Bidirectional gray matter changes after complex motor skill learning. Front Syst Neurosci. 6:37.

Hamzei F, Glauche V, Schwarzwald R, May A. 2012. Dynamic gray matter changes within cortex and striatum after short motor skill training are associated with their increased functional interaction. Neuroimage. 59:3364–3372.

Han Y, Yang H, Lv Y-T, Zhu C-Z, He Y, Tang H-H, Gong Q-Y, Luo Y-J, Zang Y-F, Dong Q. 2009. Gray matter density and white matter integrity in pianists' brain: a combined structural and diffusion tensor MRI study. Neurosci Lett. 459:3–6.

Harrington GS, Farias D, Davis CH, Buonocore MH. 2007. Comparison of the neural basis for imagined writing and drawing. Hum Brain Mapp. 28:450–459.

Hazeltine E, Grafton ST, Ivry R. 1997. Attention and stimulus characteristics determine the locus of motor-sequence encoding: A PET study. Brain. 120 (Pt 1):123–140.

Hikosaka O, Nakamura K, Sakai K, Nakahara H. 2002. Central mechanisms of motor skill learning. Curr Opin Neurobiol. 12:217–222.

Hübener M, Bonhoeffer T. 2014. Neuronal plasticity: Beyond the critical period. Cell. 159:727–737.

Ilg R, Wohlschläger AM, Gaser C, Liebau Y, Dauner R, Wöller A, Zimmer C, Zihl J, Mühlau M. 2008. Gray matter increase induced by practice correlates with task-specific activation: a combined functional and morphometric magnetic resonance imaging study. J Neurosci. 28:4210–4215.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. 2012. Fsl. Neuroimage. 62:782–790.

Katanoda K, Yoshikawa K, Sugishita M. 2001. A functional MRI study on the neural substrates for writing. Hum Brain Mapp. 13:34–42.

Kim SG, Ashe J, Hendrich K, Ellermann JM, Merkle H, Uğurbil K, Georgopoulos AP. 1993. Functional magnetic resonance imaging of motor cortex: hemispheric asymmetry and handedness. Science (80). 261:615–617.

Koppelstaetter F, Poeppel TD, Siedentopf CM, Ischebeck A, Kolbitsch C, Mottaghy FM, Felber SR, Jaschke WR, Krause BJ. 2010. Caffeine and cognition in functional magnetic resonance imaging. J Alzheimer's Dis. 20(Suppl. 1):S71–S84.

Kronenberg G, Bick-Sander A, Bunk E, Wolf C, Ehninger D, Kempermann G. 2006. Physical exercise prevents age-related decline in precursor cell activity in the mouse dentate gyrus. Neurobiol Aging. 27:1505–1513.

Kühn S, Gleich T, Lorenz RC, Lindenberger U, Gallinat J. 2014. Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. Mol Psychiatry. 19:265–271.

Kwok V, Niu Z, Kay P, Zhou K, Mo L, Jin Z, So K-F, Tan LH. 2011. Learning new color names produces rapid increase in gray matter in the intact adult human cortex. Proc Natl Acad Sci USA. 108:6686–6688.

Li A, Yetkin FZ, Cox R, Haughton VM. 1996. Ipsilateral hemisphere activation during motor and sensory tasks. Am J Neuroradiol. 17:651–655.

Lisofsky N, Mårtensson J, Eckert A, Lindenberger U, Gallinat J, Kühn S. 2015. Hippocampal volume and functional connectivity changes during the female menstrual cycle. Neuroimage. 118:154–162.

Loftus GR, Masson ME. 1994. Using confidence intervals in within-subject designs. Psychon Bull Rev. 1:476–490.

Lövdén M, Bäckman L, Lindenberger U, Schaefer S, Schmiedek F. 2010. A theoretical framework for the study of adult cognitive plasticity. Psychol Bull. 136:659–676.

Lövdén M, Schaefer S, Noack H, Bodammer NC, Kühn S, Heinze H-J, Düzel E, Bäckman L, Lindenberger U. 2012. Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. Neurobiol Aging. 33:620.e9–e620.e22.

Lövdén M, Wenger E, Mårtensson J, Lindenberger U, Bäckman L. 2013. Structural brain plasticity in adult learning and development. Neurosci Biobehav Rev. 37:2296–2310.

Luft AR, Buitrago MM. 2005. Stages of motor skill learning. Mol Neurobiol. 32:205–216.

Marquardt C, Gentz W, Mai N. 1999. Visual control of automated handwriting movements. Exp Brain Res. 128:224–228.

Mårtensson J, Eriksson J, Bodammer NC, Lindgren M, Johansson M, Nyberg L, Lövdén M. 2012. Growth of language-related brain areas after foreign language learning. Neuroimage. 63:240–244.

May A. 2011. Experience-dependent structural plasticity in the adult human brain. Trends Cogn Sci. 15:475–482.

Molina-Luna K, Hertler B, Buitrago MM, Luft AR. 2008. Motor learning transiently changes cortical somatotopy. Neuroimage. 40:1748–1754.

Morey RD. 2008. Confidence intervals from normalized data: a correction to Cousineau (2005). Tutor Quant Methods Psychol. 4:61–64.

Mottolese C, Richard N, Harquel S, Szathmari A, Sirigu A, Desmurget M. 2013. Mapping motor representations in the human cerebellum. Brain. 136:330–342.

Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9:97–113.

Overall RW, Walker TL, Leiter O, Lenke S, Ruhwald S, Kempermann G. 2013. Delayed and transient increase of adult hippocampal neurogenesis by physical exercise in DBA/2 mice. PLoS One. 8:1–12.

Park IS, Lee KJ, Han JW, Lee NJ, Lee WT, Park KA, Rhyu IJ. 2009. Experience-dependent plasticity of cerebellar vermis in basketball players. Cerebellum. 8:334–339.

Penhune VB, Doyon J. 2005. Cerebellum and M1 interaction during early learning of timed motor sequences. Neuroimage. 26:801–812.

Pletzer B, Kronbichler M, Aichhorn M, Bergmann J, Ladurner G, Kerschbaum HH. 2010. Menstrual cycle and hormonal

contraceptive use modulate human brain structure. Brain Res. 1348:55–62.

Quallo MM, Price CJ, Ueno K, Asamizuya T, Cheng K, Lemon RN, Iriki A. 2009. Gray and white matter changes associated with tool-use learning in macaque monkeys. Proc Natl Acad Sci USA. 106:18379–18384.

Reed A, Riley J, Carraway R, Carrasco A, Perez C, Jakkamsetti V, Kilgard MP. 2011. Cortical map plasticity improves learning but is not necessary for improved performance. Neuron. 70:121–131.

Romero DH, Van Gemmert AWA, Adler CH, Bekkering H, Stelmach GE. 2003. Time delays prior to movement alter the drawing kinematics of elderly adults. Hum Mov Sci. 22:207–220.

Sampaio-Baptista C, Scholz J, Jenkinson M, Thomas AG, Filippini N, Smit G, Douaud G, Johansen-Berg H. 2014. Gray matter volume is associated with rate of subsequent skill learning after a long term training intervention. Neuroimage. 96:158–166.

Sanes JN, Donoghue JP. 2000. Plasticity and primary motor cortex. Annu Rev Neurosci. 23:393–415.

Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, Phillips C, Richiardi J, Mourão-Miranda J. 2013. PRoNTo: Pattern recognition for neuroimaging toolbox. Neuroinformatics. 11:319–337.

Silver M, Montana G, Nichols TE. 2011. False positives in neuroimaging genetics using voxel-based morphometry data. Neuroimage. 54:992–1000.

Tam F, Churchill NW, Strother SC, Graham SJ. 2011. A new tablet for writing and drawing during functional MRI. Hum Brain Mapp. 32:240–248.

Taubert M, Draganski B, Anwander A, Müller K, Horstmann A, Villringer A, Ragert P. 2010. Dynamic properties of human brain structure: learning-related changes in cortical areas and associated fiber connections. J Neurosci. 30:11670–11677.

Thomas C, Baker CI. 2013. Teaching an adult brain new tricks: a critical review of evidence for training-dependent structural plasticity in humans. Neuroimage. 73:225–236.

Ungerleider LG, Doyon J, Karni A. 2002. Imaging brain plasticity during motor skill learning. Neurobiol Learn Mem. 78:553–564.

Vidyasagar R, Greyling A, Draijer R, Corfield DR, Parkes LM. 2013. The effect of black tea and caffeine on regional cerebral blood flow measured with arterial spin labeling. J Cereb Blood Flow Metab. 33:963–968.

Wenger E, Schaefer S, Noack H, Kühn S, Mårtensson J, Heinze H-J, Düzel E, Bäckman L, Lindenberger U, Lövdén M. 2012. Cortical thickness changes following spatial navigation training in adulthood and aging. Neuroimage. 59:3389–3397.

Willingham DB. 1998. A neuropsychological theory of motor skill learning. Psychol Rev. 105:558–584.

Woollett K, Maguire EA. 2011. Acquiring "the Knowledge" of London's layout drives structural brain changes. Curr Biol. 21:2109–2114.

Wrigley PJ, Gustin SM, Macey PM, Nash PG, Gandevia SC, Macefield VG, Siddall PJ, Henderson LA. 2009. Anatomical changes in human motor cortex and motor pathways following complete thoracic spinal cord injury. Cereb Cortex. 19:224–232.

Xu T, Yu X, Perlik AJ, Tobin WF, Zweig JA, Tennant K, Jones T, Zuo Y. 2009. Rapid formation and selective stabilization of synapses for enduring motor memories. Nature. 462:915–919.

Yousry TA, Schmid UD, Alkadhi H, Schmidt D, Peraud A, Buettner A, Winkler P. 1997. Localization of the motor hand area to a knob on the precentral gyrus: a new landmark. Brain. 120:141–157.

Zatorre RJ, Fields RD, Johansen-Berg H. 2012. Plasticity in gray and white: neuroimaging changes in brain structure during learning. Nat Neurosci. 15:528–536.