Supplementary Methods.

Inclusion criteria. Inclusion criteria for all subjects were: a) right-handedness; b) normal/corrected-to-normal vision; c) no neurological (apart MS), orthopedic or rheumatologic disorders, d) no pain disturbances; e) no psychiatric/mood disorders; f) no concomitant antidepressants, baclofen, psychoactive, or steroid therapy as well as no treatment for symptoms of fatigue; g) no drug/alcohol abuse; and h) no particular manual skills (i.e., musicians, athletes, typewriting). Additional inclusion criteria for MS patients were: a) no relapses and steroid treatment in the three months preceding study initiation; b) stable treatment for MS for at least 6 months; and c) no rehabilitation during the previous three months.

Upper limb training protocol. For upper limb rehabilitation, all subjects were randomly assigned, through a computer-generated sequence, to an AOT group (experimental group) and a control group, who underwent two different schemes of motor training during the study period. The training lasted two weeks and consisted of ten sessions (during week days) of forty minutes each, which included ten minutes of passive mobilization of the right upper limb, viewing of three videos (which were different for the AOT and control groups) and execution, with the right hand, of three daily-life actions shown in the videos seen by the AOT group. During the training sessions, all subjects sat relaxed on a chair with their arms placed on a table. Participants had a free view of a large computer screen (25 inches), positioned one meter in front of them. Subjects in the AOT group were asked to carefully watch video sequences lasting five minutes containing daily-life right hand and arm actions, and then to repeatedly practice the observed actions for five minutes with their right upper limb, using the same objects shown in the video. In each session, three videos were watched, alternated by the performance of the three corresponding actions. Control group subjects underwent the same procedure as the AOT group with the exception that they watched videos of inanimate landscapes, which are unlikely to elicit activity in motor or MNS areas. The practiced right hand and arm actions were performed after instruction by the assisting physiotherapist in the

exact order as they were practiced by the AOT group, so that the "active" intervention was equivalent across groups.

By the end of the ten training sessions, each subject had been trained with 30 different daily-life actions of increasing complexity. In order to standardize treatment, but make it as individualized as possible, these 30 gestures were chosen from a set of 80 different daily-life actions of increasing difficulty. HC performed the most difficult 30 gestures, while after the baseline assessment an expert physiotherapist, blind to treatment, chose the 30 actions to be performed by each MS patient based on the level of individual motor impairment. The actions shown during the videos were performed consecutively by an actor and an actress, who performed the action both in first and third person. The actors used only their right hand and common objects belonging to the everyday life environment.

MRI acquisition. In all subjects, before and after the training, using a 3.0 Tesla Philips Intera scanner (Philips Medical Systems, Eindhoven, The Netherlands), the following brain scans were acquired: a) T2*-weighted single-shot EPI sequence for resting state (RS) fMRI (TR=3000 ms, TE=35 ms, flip angle=90°, FOV=240x240mm^2; matrix=128x128, 200 sets of 30 contiguous axial slices, with a thickness of 4 mm); b) T2*-weighted single-shot EPI sequence during a manipulation task (TR=3700 ms, TE=35 ms, flip angle=85°, FOV=240mm^2; matrix=128x128, 60 sets of 30 contiguous axial slices, with a thickness of 4 mm); c) dual-echo turbo spin echo (TSE) (TR/TE=2599/16-80 ms; flip angle=90°; FOV=240mm^2; matrix=256×256; ETL=6; 44 contiguous, 3-mm-thick axial slices); d) 3D T1-weighted spoiled transient field echo (TFE) (TR/TE/TI=7/3.5/1000 ms; flip angle=8°; matrix size=256x256; FOV=256x256 mm^2; 150 contiguous, axial slices with voxel size=1x1x1 mm); and e) pulsed-gradient SE EP (TE/TR=58/8938 ms, acquisition matrix size=112x88, FOV=240x231 mm^2, 55 contiguous, 2.3-mm thick axial slices) with SENSE (acceleration factor=2) and diffusion gradients applied in 35 non-collinear directions (two optimized *b* factors were used: b_1=0, b_2=900 s/mm^2).

For all scans, the slices were positioned to run parallel to a line that joins the most infero-anterior and infero-posterior parts of the corpus callosum.

Total acquisition time of the RS fMRI sequence was 10 minutes. During RS fMRI scanning, subjects were instructed to keep their eyes closed, to remain motionless and not to think anything in particular. All subjects reported that they had not fallen asleep during scanning, according to a questionnaire delivered immediately after the MRI session. During the manipulation task, subjects laid supine in the scanner, with their eyes closed. Using a block design (ABAB), where six periods of the active condition were alternated with six periods of the baseline condition (each period consisting of five measurements), all subjects were asked to manipulate six different complex common objects of a size fitting easily into the palm (a pencil, a toothbrush, a spoon, a syringe, a glass and a fork) alternating with a neutral one (a sphere). Objects were passed to the subject by an experimenter inside the scanner room. Common objects were randomly selected and not re-used during a session. Two sessions were conducted, one for each hand. All MS patients were able to maintain the object in their affected hand and manipulate it in the correct way.

Voxel-based morphometry (VBM) and tensor-based morphometry (TBM). TBM, as implemented in SPM12 (www.fil.ion.ucl.ac.uk/spm), was used to map changes in regional GM volumes over time in the four study groups. Pairwise longitudinal registration was used to align the first and second scan of each subject.[1] The method is based on pairwise inverse-consistent registration and incorporates a bias field correction. Prior to running the non-linear registration, each subject's baseline scan was coregistered to the follow up scan using a rigid-body transformation and *vice versa*. The absolute values of the two transforms were averaged, halved and written in the image headers with opposite signs from each other, to obtain a halfway coregistration without image reslicing. The rate of volume change was quantified by saving the velocity field divergence map, where positive values indicate expansion and negative values contraction. The mid-point average template image was also saved. This was used for groupwise alignment: first, the mid-point average template images were segmented into different tissue types via the Segmentation

routine in SPM12.[2] Then, GM and WM segmented images of all subjects, in the closest possible rigid-body alignment with each other, were used to produce GM and WM templates and to drive the deformation to the templates. At each iteration, the deformations, calculated using the Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL) registration method,[3] were applied to GM and WM, with an increasingly good alignment of subject morphology, to produce templates. Finally, an affine transformation that maps from the population average (DARTEL Template space) to MNI space was calculated, divergence maps were spatially normalized and smoothed with an 8 mm Gaussian kernel. These last 3 steps are incorporated in a unique tool, called "Normalise to MNI Space". The steps described for groupwise alignment were repeated for baseline 3D TFE images to run a VBM analysis. The only difference in the procedure described above is that the normalization to MNI space was applied to GM maps, and that, after transformation these were modulated to ensure that the overall amount in each tissue class was not altered by the spatial normalisation procedure. Regional differences in GM volume at baseline and the rate of volume change over the follow up were assessed using the general linear model and the theory of Gaussian fields.[4] To limit the analysis to the GM, an inclusion mask obtained from the GM DARTEL Template, transformed to the MNI space, smoothed and thresholded at 0.5, was used. VBM analysis was also corrected for intracranial volume. VBM and TBM results were assessed at a threshold of $p<0.05$, family-wise error corrected (FWE) for multiple comparison and also tested at a $p<0.001$, uncorrected (kE=5). For between-group comparisons, a $p<0.05$ clusterwise FWE corrected threshold was applied.[5]

TBSS analysis. Diffusion-weighted images were corrected for distortions induced by the eddy currents and for head movements, and transformed to MNI space (http://white.stanford.edu/mrdiff). Then, using the FMRIB's Diffusion Toolbox (FDT tool, FSL 4.1, http://www.fmrib.ox.ac.uk), the diffusion tensor (DT) was estimated in each voxel by linear regression[6] and mean diffusivity (MD), radial diffusivity (RD), axial diffusivity (AD) and fractional anisotropy (FA) maps derived. TBSS analysis was used for voxelwise analysis of

whole brain WM DT-MRI measures (http://www.fmrib.ox.ac.uk/fsl/tbss/index.html). In detail, individual FA images (both baseline and follow up) were non-linearly registered to the FMRIB58_FA atlas[7] provided within FSL, and averaged. The resulting mean FA image was thinned to create a WM tract "skeleton", which was thresholded at a FA > 0.2 to include only WM voxels. Individual subjects' FA values were projected onto this group skeleton by searching perpendicular from the skeleton for maximum FA values. Maximum FA values were chosen in order to restrict analysis to the centers of WM tracts (where maximum FA values are found), rather than considering voxels at the edge of tracts, that may suffer from partial volume effects. The individual registration and projection vectors obtained during this process were also applied to the RD, AD, and MD data.

Voxelwise between-group differences of DT MRI metrics at baseline and longitudinal changes were tested using a permutation method ("Randomise" program within FSL), including age and sex as covariates. The number of permutations was set at 5000. A p value<0.05 (FWE corrected) and the threshold-free cluster enhancement option[8] in the Randomize tool were used.

Active and RS fMRI pre-processing. Using SPM12, active and RS fMRI data were realigned to the mean image from all images in the series, to correct for subject motion with a six degrees of freedom rigid-body transformation and co-registered to the 3D T1-weighted anatomical image of the same subject. Data were spatially normalized into the standard MNI space by applying the non-linear warping parameters used to transform the 3D T1-weighted scan to the MNI standard space. Then, using the REST software (http://resting-fmri.sourceforge.net), RS fMRI data were linearly detrended and band-pass filtered between 0.01 and 0.08 Hz to partially remove low-frequency drifts and physiological high-frequency noise. Non-neuronal sources of synchrony within the RS fMRI time series were removed by regressing out the six motion parameters estimated by SPM12 and the average signals from the ventricular CSF and WM (segmented on the 3D T1-weighted scan of each study subject using SPM12). Finally, active and RS fMRI data were smoothed using a 3D Gaussian filter (full-width at half maximum = 10 mm and 6 mm, respectively).

Active fMRI analysis. Changes in BOLD contrast associated with the performance of the manipulation task were assessed using the general linear model and the theory of Gaussian fields. All subjects were included in the subsequent statistical analysis because all of them had a maximum translation/rotation that gave movements lower than 3.0 mm in the x, y, and z directions. The motion parameters derived from realignment were included in the general linear model as confounding regressors. Specific effects were tested by applying appropriate linear contrasts. Significant hemodynamic changes for each contrast were assessed using t statistical parametric maps (SPMt).

RS FC analysis. Since the estimation of the hand motor network is challenging, with spatial extent and location of this network being highly variable across subjects,[9] we performed a preliminary ICA to detect a component related to the functional sensorimotor system, which was subsequently used as a constraint for the exact location of the left hand motor seed, as described in detail below. ICA was run using the GIFT software package (http://mialab.mrn.org/software/gift/). The component with the highest correlation coefficient (R2, as implemented in GIFT) with a template of the sensorimotor system[10] was selected for this purpose.

A seed-based correlation approach was used to assess RS FC modifications within the MNS (using the left [L] inferior frontal gyrus [IFG] as a seed) and the hand motor network (using the L hand motor cortex as a seed).[11] Masks of the L IFG were obtained by merging the masks of L Brodmann Areas (BA) 44 and 45 included in the WFU PickAtlas toolbox (http://fmri.wfubmc.edu/software/PickAtlas). The seed for the hand motor network was a sphere of 6-mm radius centered in the L BA4. The exact location of the seed center was determined for each subject by using the following procedure: 1) manual assessment of the anatomical coordinates of the L hand motor cortex by the omega shape on the 3D T1-weighted scans (co-registered on RS fMRI images), read by one experienced neurologist; 2) shift of the omega shape-centered seed to match the closest peak of the subject IC sensorimotor component estimated by GIFT, constrained to be within the L BA4. After the creation of seed regions, RS FC was investigated by calculating the

correlation coefficients between the time series extracted from L IFG and L hand motor region and any other voxel in the brain. A Fisher's z transformation was used to improve the Gaussianity of the correlation coefficients obtained.[12]

Intra-group activations/RS FC and comparisons between groups were investigated using linear models, adjusted for age and sex.[13] All fMRI results were assessed at a threshold of $p<0.05$, FWE corrected for multiple comparisons and also tested at a $p<0.001$, uncorrected (kE=5). For between-group comparisons, a $p<0.05$ clusterwise FWE corrected threshold was applied.[5] Multiple linear regression models, adjusted for age and sex, were used to assess correlations between fMRI measures and functional tests ($p<0.001$, uncorrected, kE=5).

Supplementary Table. Results of between-group longitudinal analysis of regional GM volumes, fMRI activations and RS FC of the motor network (using the left hand motor cortex as a seed) and the mirror neuron system network (using the left inferior frontal gyrus as a seed) (random-effect analysis and full factorial models, $p<0.001$, uncorrected. Clusters surviving at $p<0.05$, family-wise corrected for multiple comparisons, are marked with *).

Analysis	Groups	Variation	Side	Brain region	MNI coordinates (x, y, z)	t value
GM volume	HC-AOT *vs* HC-Control	Decrease	R	Lingual gyrus	12, -72, -2	4.89*
		Increase	L	Superior frontal gyrus	-24, 63, 9	5.56*
	HC-Control *vs* HC-AOT	Increase	L	Inferior temporal gyrus	-54, -15, -16	5.71*
			R	Middle frontal gyrus	40, 44, 10	4.91*
	MS-AOT *vs* MS-Control	Decrease	R	Supplementary motor area	6, -8, 62	5.92*
		Increase	L	Middle occipital gyrus	-30, -81, 2	6.01*
			R	Inferior frontal gyrus	10, 45, -3	5.47*
			L	Superior frontal gyrus	-3, 45, 39	5.49*
	MS-Control *vs* MS-AOT	Decrease	R	Middle temporal gyrus	46, -57, 2	6.19*
	HC-AOT *vs* MS-AOT	Increase	R	Middle occipital gyrus	27, -98, 12	3.35
		Decrease	R	Superior temporal gyrus	54, -15, -4	3.65
			L	Inferior temporal gyrus	-54, -46, -10	3.54
			L	Superior parietal lobule	-26, -58, 45	3.51
	MS-AOT *vs* HC-AOT	Increase	L	Angular	-42, -60, 38	3.68
	HC-Control *vs* MS-Control	Increase	R	Middle frontal gyrus	28, 22, 48	4.08
			L	Posterior cingulum	-8, -44, 21	3.28
		Decrease	L	Parahippocampal gyrus	-26, -30, -18	3.84
			R	Middle occipital gyrus	33, -78, 30	3.54

R Manipulation	HC-AOT *vs* HC-Control	Increase	L	Precentral gyrus	-48, 2, 34	5.59*
	MS-AOT *vs* MS-Control	Increase	L	Inferior frontal gyrus	-54, 32, 8	6.12*
			L	Insula	-30, 12, -16	5.94*
			R	Inferior frontal gyrus	40, 24, -10	5.93*
	MS-Control *vs* HC-Control	Decrease	L	Superior termporal gyrus	-64, -22, 14	3.73
L Manipulation	MS-AOT *vs* MS-Control	Increase	L	Fusiform gyrus	-34, -20, -22	6.02*
			R	Superior frontal gyrus	20, 62, 10	5.73*
	MS-AOT *vs* HC-AOT	Increase	R	Parahippocampal gyrus	34, -30, -16	3.57
	MS-Control *vs* HC-Control	Decrease	L	Middle temporal gyrus	-54, -12, -10	3.57
Motor network	HC-Control *vs* HC-AOT	Increase	R	Supplementary Motor Area	12, 22, 64	6.23*
	MS-AOT *vs* MS-Control	Increase	L	Cerebellum	-6,-66, -32	6.22*
			R	Inferior frontal gyrus	50, 10, 26	4.95*
	MS-Control *vs* MS-AOT	Increase	R	Supplementary Motor Area	14, -4, 72	6.33*
	HC-Control *vs* MS-Control	Decrease	L	Cerebellum	-44 -80 -30	3.39
			R	Cerebellum	16, -82, -48	3.31
	MS-Control *vs* HC-Control	Increase	R	Supplementary Motor Area	6, -2, 70	4.64
MNS network	MS-AOT *vs* MS-Control	Increase	R	Cerebellum	16, -66, -12	6.25*
			R	Calcarine suclus	16, -100, 0	5.89*
	MS-Control *vs* MS-AOT	Increase	L	Anterior cingulum	-6, 52, 4	6.12*
	HC-AOT *vs*	Increase	R	Inferior frontal gyrus	46, -22, 12	4.24*

	MS-AOT					
	MS-AOT *vs* HC-AOT	Increase	R	Calcarine sulcus	18, -102, 0	4.72*
			L	Middle occipital gyrus	-38, -90, 8	4.13
			L	Cerebellum	-18 -58 -16	4.06
			L	Calcarine sulcus	-6, -98, 2	3.93
			R	Middle occipital gyrus	40, -86, 16	3.57
			R	Cerebellum	24, -76, -22	3.52
			R	Middle temporal gyrus	50, -68, 20	3.46
			L	Supramarginal gyrus	-62, -24, 44	3.22
	HC-Control *vs* MS-Control	Increase	R	Postcentral gyrus	26, -36, 76	3.31
	MS-Control *vs* HC-Control	Increase	L	Cerebellum	-12 -46 -38	4.87*
			L	Insula	-34, 10, 10	3.78
			L	Superior Frontal	-20, 54, 20	3.76
			L	Anterior Cingulum	-14, 32, 28	3.74

Abbreviations: GM=gray matter; MNS=mirror neuron system; HC=healthy controls; MS=Multiple Sclerosis; AOT=action observation therapy; R=right; L=left; MNI=Montreal Neurological Institute.

References

1. Ashburner J and Ridgway GR. Symmetric diffeomorphic modeling of longitudinal structural MRI. *Frontiers in neuroscience*. 2012; 6: 197.
2. Ashburner J and Friston KJ. Unified segmentation. *NeuroImage*. 2005; 26: 839-51.
3. Ashburner J. A fast diffeomorphic image registration algorithm. *NeuroImage*. 2007; 38: 95-113.
4. Friston KJ, Holmes AP, Poline JB, et al. Analysis of fMRI time-series revisited. *NeuroImage*. 1995; 2: 45-53.
5. Friston KJ, Worsley KJ, Frackowiak RS, Mazziotta JC and Evans AC. Assessing the significance of focal activations using their spatial extent. *Human brain mapping*. 1994; 1: 210-20.
6. Basser PJ, Mattiello J and LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of magnetic resonance Series B*. 1994; 103: 247-54.
7. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO and Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE transactions on medical imaging*. 1999; 18: 712-21.
8. Smith SM and Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*. 2009; 44: 83-98.
9. Hou BL, Bhatia S and Carpenter JS. Quantitative comparisons on hand motor functional areas determined by resting state and task BOLD fMRI and anatomical MRI for pre-surgical planning of patients with brain tumors. *NeuroImage Clinical*. 2016; 11: 378-87.
10. Smith SM, Fox PT, Miller KL, et al. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106: 13040-5.
11. Biswal B, Yetkin FZ, Haughton VM and Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*. 1995; 34: 537-41.
12. Lowe MJ, Mock BJ and Sorenson JA. Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. *NeuroImage*. 1998; 7: 119-32.
13. Friston KJ, Holmes AP, Price CJ, Buchel C and Worsley KJ. Multisubject fMRI studies and conjunction analyses. *NeuroImage*. 1999; 10: 385-96.