OXFORD UNIVERSITY PRESS | Cerebral Cortex

Cognitive control structures in the imitation learning of spatial sequences and rhythms – a fMRI study

Journal:	*Cerebral Cortex*
Manuscript ID	CerCor-2016-00976.R1
Manuscript Type:	Original Articles
Date Submitted by the Author:	n/a
Complete List of Authors:	Sakreida, Katrin; University Hospital RWTH Aachen, Department of Neurosurgery Higuchi, Satomi; Lancaster University, Department of Psychology; University of Liverpool, Magnetic Resonance and Image Analysis Research Centre; Hokkaidō University, Center for Experimental Research in Social Sciences Di Dio, Cinzia; Universita Cattolica del Sacro Cuore, Department of Psychology Ziessler, Michael; Liverpool Hope University, Department of Psychology Turgeon, Martine; Centre-Est-de-l'Ile-de-Montréal, Centre de recherche interdisciplinaire en réadaptation (CRIR), Centre intégré universitaire de santé et de services sociaux (CIUSSS) Roberts, Neil; University of Edinburgh, School of Clinical Sciences, Clinical Research Imaging Centre (CRIC) Vogt, Stefan; Lancaster University, Department of Psychology; University of Liverpool, Magnetic Resonance and Image Analysis Research Centre
Keywords:	cognitive control, fronto-parietal mirror circuit, motor imagery, musical expertise, performance monitoring

Cognitive control structures in the imitation learning of spatial sequences and rhythms – a fMRI study

Katrin Sakreida[1*], Satomi Higuchi[2,3,4], Cinzia Di Dio[5], Michael Ziessler[6], Martine Turgeon[7], Neil Roberts[8], Stefan Vogt[2,3*]

[1]Department of Neurosurgery, Medical Faculty, RWTH Aachen University, Aachen, Germany
[2]Department of Psychology, Lancaster University, Lancaster, United Kingdom
[3]Magnetic Resonance and Image Analysis Research Centre, University of Liverpool, Liverpool, United Kingdom
[4]Center for Experimental Research in Social Sciences, Hokkaidō University, Sapporo, Japan
[5]Department of Psychology, Universita Cattolica del Sacro Cuore, Milan, Italy
[6]Department of Psychology, Liverpool Hope University, Liverpool, United Kingdom
[7]Centre de recherche interdisciplinaire en réadaptation (CRIR), Centre intégré universitaire de santé et de services sociaux (CIUSSS) du Centre-Est-de-l'Ile-de-Montréal, Montréal, Québec, Canada
[8]Clinical Research Imaging Centre (CRIC), School of Clinical Sciences, University of Edinburgh, Edinburgh, Scotland, United Kingdom

***Corresponding authors:**
University Hospital of RWTH Aachen University
Department of Neurosurgery
Pauwelsstr. 30
52074 Aachen, Germany
ksakreida@ukaachen.de
Tel. +49-241-80 80233
Fax: +49-241-80-82420

Dr. Stefan Vogt
Department of Psychology
Lancaster University
Lancaster LA14YF, UK
s.vogt@lancaster.ac.uk
Tel. +44-1524-594625
Fax: +44-1524-593744

Running title: Cognitive control structures in imitation learning

Revised for *Cerebral Cortex*

Number of words in abstract: 193
Update number of words in main text: 10195 [originally 12670]
Update number of words for introduction: 1658 [originally 1778]
Update number of words for discussion / conclusion: 4305 [originally 5387]
Number of figures: 6 (plus 3 supplementary figures)
Update number of tables: 2 (plus 4 supplementary tables)

Abstract

Imitation learning involves the acquisition of novel motor patterns based on action observation. We used event-related functional magnetic resonance imaging to study the imitation learning of spatial sequences and rhythms during action observation, motor imagery, and imitative execution in non-musicians and musicians. Whilst both tasks engaged the fronto-parietal mirror circuit, the spatial sequence task recruited posterior parietal and dorsal premotor regions more strongly. The rhythm task involved an additional network for auditory working memory. This partial dissociation supports the concept of task-specific mirror mechanisms. Two regions of cognitive control were identified: (1) Dorsolateral prefrontal cortex (DLPFC) was found to be more strongly activated during motor imagery of novel spatial sequences, which allowed us to extend the two-level model of imitation learning by Buccino et al. (2004) to spatial sequences. (2) During imitative execution of both tasks, the posterior medial frontal cortex was robustly activated, along with the DLPFC, which suggests that both regions are involved in the cognitive control of imitation learning. The musicians' selective behavioural advantage for rhythm imitation was reflected cortically in enhanced sensory-motor processing during action observation and by the absence of practice-related activation differences in DLPFC during rhythm execution.

Keywords: cognitive control, fronto-parietal mirror circuit, motor imagery, musical expertise, performance monitoring

Introduction

Imitation learning involves the acquisition of novel motor patterns based on action observation and motor execution, and it is one of the most frequently used forms of skill acquisition in occupational, sports, musical, and rehabilitation settings. In the present study we explore the neuro-cognitive mechanisms underlying imitation learning for a prototypical task domain, namely imitation of sequences of finger movements. The central motivation for this study was to test Buccino et al.'s (2004) two-level model of imitation learning with sequential actions. This model comprises a core task network for sensorimotor encoding and the dorsolateral prefrontal cortex (DLPFC) as cognitive control hub. It has been supported in a series of functional magnetic resonance imaging (fMRI) studies (Buccino et al. 2004; Vogt et al. 2007; Higuchi et al. 2012), which used the learning of guitar chords as an example of complex skill acquisition. However, such configural actions, or bodily postures, represent just one class of motor skills (for review see Vogt and Thomaschke 2007). With the present work we were therefore seeking to establish if Buccino et al.'s model can be extended to sequence learning.

We pursued three main research objectives: (1) to delineate the core task networks for two different forms of motor sequencing, namely sequences of spatially oriented finger movements (SEQ) and rhythmical sequences (RHY), (2a) to describe the functional reorganisation in both task networks after a moderate amount of practice as well as (2b) at different levels of expertise, and, crucially, (3) to explore, on this basis, the involvement of cognitive control structures, including the DLPFC, in the early stages of sequence learning. Here we were interested (3a) in the specific cognitive control structures involved in the two tasks and (3b) in task-specific expertise effects. To this end, we studied both musically naïve and expert participants. The latter group generally exhibits advanced capabilities of encoding

rhythmical patterns (Matthews et al. 2016), whilst for the spatial sequences we expected (and found) similar levels of performance in both groups. In the SEQ task, participants observed and then imitated an index finger pressing a series of eight keys on a four-key keyboard, and in the RHY task, they imitated the same finger producing a series of eight intervals on the same key with a mix of long, medium, and short durations. Half of these patterns had been practised one day before the scanning, the other half was novel.

The available neuroimaging literature on imitation learning is remarkably sparse. However, two clusters of research are directly relevant to the present study, first the extensive neuroimaging work on action observation and on the imitation of familiar actions ('familiar imitation', Subiaul 2010), and second the neuroimaging literature on the acquisition, consolidation, and retention of motor skills, where a good part of this literature concerns motor sequencing. In the following, we develop the predictions regarding the three research objectives from key findings in these two research areas.

From action observation and familiar imitation to imitation learning. There is substantial evidence that observing the actions of others can induce processing in motor cortical regions of the observer's brain (Rizzolatti et al. 2014; see also meta-analyses by Caspers et al. 2010, and Molenberghs et al. 2012). A plausible general account is that this motor cortical 'mirroring' is part of a generative model that predicts the sensory input (Kilner et al. 2007; Kilner and Lemon 2013). When imitating familiar actions (or 'behavioural mimicry', Chartrand and van Baaren 2009), this generative model can also be used to guide motor execution of the observed behaviour (Vogt 2002; Caspers et al. 2010).

In contrast to familiar imitation, imitation learning requires the generation of novel behaviour which is not readily available in the observer's motor repertoire. In the first neuroimaging study on this topic, Buccino et al. (2004) found that the classic regions of the

human fronto-parietal mirror circuit, namely ventral premotor cortex (PMv), pars opercularis of the inferior frontal gyrus (IFG), and inferior parietal lobule (IPL), were strongly activated from the very outset of imitation learning. Most likely, this reflects the segmentation of the observed action into its constituent elements (e.g., individual fingers), which would normally be present in the observer's motor repertoire (Byrne 2003; Rizzolatti 2014). Whilst the majority of studies on action observation have focused on prehensile actions, recent research indicates that the task networks for action observation can substantially vary with the nature of the task. Regarding the task networks subserving the present SEQ and RHY tasks, we expected areas of overlap in the fronto-parietal mirror circuit (Caspers et al. 2010; Konoike et al. 2012), and the supplementary motor area (SMA, Vogt et al. 2007; Mukamel et al. 2010; Dayan and Cohen 2011; Hardwick et al. 2013), as well as task-specific differences (*research objective 1*). Regarding the latter, we expected a stronger involvement of posterior parietal regions for the SEQ task than for the RHY task, and the recruitment of additional brain regions for encoding temporal information in the RHY task. Such dissociations between the present, visually well-matched SEQ and RHY tasks would directly support the concept of task-specific mirror mechanisms (Subiaul 2010; Rizzolatti et al. 2014).

In addition to the core fronto-parietal mirror circuit, Buccino et al. (2004) found the DLPFC activated during motor preparation of imitative execution. In a follow-up study (Vogt et al. 2007), the DLPFC was more strongly involved during observation and preparation of novel hand postures, compared to previously practised hand postures. Using a rapid imitation task Higuchi et al. (2012) confirmed the latter finding for imitative execution and demonstrated a robust connectivity between left DLPFC and the fronto-parietal mirror circuit. In addition, the behavioural benefit of imitation learning was significantly correlated with prefrontal activation intensities during observation of novel actions. Taken together, this set of

results provides compelling evidence for a crucial role of prefrontal cortex in the early stage of imitation learning. We concluded that the visuo-motor representation of an observed action, as provided by the fronto-parietal mirror circuit, "only serves as the 'raw material' for higher-order supervisory and monitoring operations associated with the prefrontal cortex" (Higuchi et al. 2012, p. 1668; Rizzolatti 2014). A structurally similar two-level model of imitation control was recently proposed by Wang and Hamilton (2012; see also Hamilton 2015), with reference to findings indicating the involvement of medial prefrontal cortex in the inhibition and selection of imitative behaviour based on social context. As already indicated, the core objective of the present study is to delineate the cognitive control hubs involved in the imitation learning of sequencing tasks. In addition to action observation (AO) and imitative execution (EXE) we also used a motor imagery (MI) condition, which replaced the motor preparatory event in our earlier studies.

From motor skill learning to imitation learning. Motor sequencing is one of the best studied task domains in the neuroimaging literature on skill learning (Doyon and Benali 2005; Dayan and Cohen 2011). There are now detailed accounts of 'fast' *versus* 'slow' motor learning and of the plastic redistribution of activations associated with each timescale (see also Kelly and Garavan 2005; Lohse et al. 2014). In keeping with our earlier work (Buccino et al. 2004; Vogt et al. 2007; Higuchi et al. 2012) the focus of the present study is on the initial stage of imitative skill learning, that is, the very first attempts at imitating a given action. Curiously, this aspect of sequence learning has been neglected in mainstream neuroimaging research. One reason for this is that research has focussed on the distinction between explicit and implicit sequence learning, with the widespread use of Nissen and Bullemer's (1987) serial reaction time (SRT) task. Here participants respond, keypress by keypress, to individual location or colour stimuli. This procedure does not represent the more typical everyday

scenario where at first a whole melody, phrase, or rhythm is attended to, before this is reproduced as a whole. Our tasks resemble this scenario. In contrast, the majority of neuroimaging studies on explicit sequence learning either used variants of the SRT task, or where this was not the case, the to-be-learned sequences were often taught informally outside the scanner (Lohse et al. 2014).

For deriving predictions regarding the to-be-expected practice effects in the present study (*research objective 2*), the following general trends observed for fast motor skill learning are relevant (Dayan and Cohen 2011): (1) the initial activation of high-level 'scaffolding' areas such as the DLPFC involved in cognitive control (Petersen et al. 1998; Shallice et al. 2004), associated with (2) the early upregulation of information processing in task-related sensory-motor regions, or task networks (Kelly and Garavan 2005; Halsband and Lange 2006), and (3) a subsequent trend towards 'neural efficiency' (see also Babiloni et al. 2009, 2010), that is, decreases in the extent and intensity of activations in cognitive control structures as well as in most, but not all components of the relevant task network. Since we had observed exactly these trends previously in action observation, motor execution, or both (Vogt et al. 2007; Higuchi et al. 2012), we expected the same overall trends in the present study. Two qualifications, however, are worth flagging here: First, Robertson et al. (2001) found that disruption of DLPFC prevented implicit sequence learning when this was guided by spatial cues, but not with guidance by colour cues. Given that spatial information was only critical in our SEQ task, it is then conceivable that the RHY task might rely less on cognitive control by the DLPFC. Second, in their recent network-analysis of explicit learning of complex, ten-element sequences, Bassett et al. (2015), found, in line with Petersen et al.'s (1998) scaffolding-storage framework, an increasing autonomy of sensorimotor systems along with a "release of cognitive control hubs" in frontal and cingulate cortices, where both regions

predicted individual differences in learning. For the present study, we were thus open-minded regarding the involvement of frontal regions other than DLPFC, and notably the posterior medial frontal cortex (pMFC), given its prominent role in performance monitoring (Ridderinkhof et al. 2004; Ullsperger et al. 2014).

Materials and Methods

Participants

Sixteen volunteers without musical experience (nine female, seven male, age range 18–23 years, mean age 20.4 ± 1.5 years) and 15 musicians (seven female, eight male, age range 18–25 years, mean age 20.8 ± 2.3 years) participated in the study. None of them had any MRI specific contraindications, or any history of neurological or psychiatric disposition.

The data of three musically naïve participants were excluded from the fMRI analysis: Two participants showed excessively large head movement during scanning, whereby the degree of movement exceeded the image voxel size, and one participant showed exceptionally poor performance for the practised patterns during scanning. Thus, the analysis comprised data of 13 participants without musical experience, and all 15 musicians. Another two musically naïve volunteers were excluded from the outset since they showed poor rhythm imitation skills in an initial screening.

Written informed consent was obtained from all participants. All had normal or corrected-to-normal visual acuity, and were strongly to moderately right-handed (mean Laterality Quotient for the non-musicians 96.9, and for the musicians 82.7) according to the Edinburgh Handedness Inventory (Oldfield 1971). Two of the musicians were ambidextrous.

The experimental procedures were approved by the local ethics committee. Data were handled anonymously, and participants were paid to compensate for their time.

The non-musicians were primarily students at the University of Liverpool. The inclusion criterion was that they should not have played any musical instrument in the last five years prior to the experiment, and have less than three years of musical experience in total. The musicians were recruited from the Liverpool Institute of Performing Arts, and from the Music department at the University of Liverpool. They had been practising the following musical instruments for 11.6 ± 3.4 years overall: guitar (n = 4), drums/percussion (n = 3), voice (n = 3), cello, flute, oboe, piano, and saxophone (n = 1 each). At the time of testing the musicians were practising their instruments on 5.1 ± 1.8 days per week for approx. 10.9 hours.

Stimuli and apparatus

Presentation software (NeuroBehavioral Systems, Berkeley, CA, USA, Version 10.1) was used for display of the stimuli and collection of responses on a custom-made four-key keyboard (see Figure 1). A total of four sets of three spatial sequences (SEQ), and four sets of three rhythms (RHY) were used, where each participant was assigned one SEQ set and one RHY set as practice sets. The to-be-practised and non-practised stimulus sets were counterbalanced across participants. The stimuli were soundless video clips of 4.7s duration, showing a right index finger performing either a SEQ or a RHY pattern on the same keyboard that was used for collecting the responses in the scanner. In each clip, the index finger started moving from a centre position between the second and third key. The SEQ stimuli consisted of eight keypresses with a fixed interval of 500 ms between keypresses. After each of the four keys was pressed once in a certain order, each key was pressed again in a different order, and the same key was never used twice in a row. For the RHY stimuli, only the third key (from left,

see Figure 1) was used, where the index finger tapped eight time intervals in a given order, comprising one long interval (L, 1000 ms), three medium intervals (M, 500 ms), and four short intervals (S, 250 ms). For instance, a spatial sequence comprised keys 1, 4, 3, 2, 3, 2, 1, 4, and a rhythm comprised the intervals M, S, S, M, L, M, S, S.

In order to ensure the comparability of performance levels in the SEQ and RHY tasks, patterns of similar difficulty were selected on the basis of a pilot study with twelve musically naïve participants, comprising a larger set of stimuli than required for the actual experiment.

Design and procedure

All participants attended a practice session outside the MRI scanner, followed by the main scanning session one day thereafter. This procedure (e.g., Vogt et al. 2007; Higuchi et al. 2012) allowed us to directly contrast patterns which had been previously practised with non-practised patterns. In the scanning session, we used a 3 x 2 x 2 experimental design (AO / MI / EXE; SEQ / RHY; practised / non-practised; see section 'Scanning session' below).

Practice session

In this session each participant was given extensive practice with one SEQ set and one RHY set in a separate room. In order to accustom participants to the scanner setup, they were lying on a bed, and stimuli were presented on a 15 inch display that was mounted approximately 75 cm above their head. Participants used their left index finger for imitation on a similar keyboard as that shown in the videos and were instructed to imitate each pattern as a mirror image of the observed pattern. This spatial arrangement preserved the spatial compatibility between display and imitation (e.g., Koski et al. 2003).

The practice session began with repeated imitation of each of the six to-be-practised patterns until this was correctly imitated over three consecutive trials. Each trial involved observation followed by execution. In order to enhance imitation accuracy, this procedure was repeated with the addition that participants were asked to perform each pattern in synchrony with the model. The second part of the practice session comprised imitation of the six to-be-practised patterns in random order for 2 x 24 trials, as well as six free recall trials. Throughout the experiment participants were discouraged from using counting or verbal labels to encode the stimuli. Finally, participants were introduced to motor imagery (MI) trials, which involved imagining the just observed sequence or rhythm and how it would feel to perform it (for further details on motor imagery see Vogt et al. 2013). They were then given a mix of trials comprising motor imagery and imitative execution of the practised patterns. In a last practice block, non-practised patterns were added so that participants experienced a similar trial composition as in the scanning session on the following day. Overall, each of the six to-be-practised patterns was imitated approx. 27 times (15 times on average in the initial imitation blocks, nine times in the trials with random order, and three times in the final set of MI and execution trials).

Scanning session

Before entering the scanning room, participants received a short booster session in the practice room, where they imitated the six practised patterns in random order for approx. 6 min and then received a short run with the same trial composition as in the scanning sessions. During scanning, participants were positioned supine with their left index positioned on the custom-made keyboard. Form-fitting cushions were used to prevent arm, hand, and head motion. Participants were provided with earplugs to attenuate scanner noise. Visual stimuli were

displayed by a LCD data projector (Panasonic PT-L785U) onto a rear-projection screen at the head end of the scanner. Participants could watch this screen via a mirror above their head. They did not see their hand during scanning. In addition to the logging of key presses via Presentation software, participants' hand movements were videotaped on MiniDV cassettes, together with an image of the displayed stimuli. In preparation of the functional analysis, the videos served the elimination of events in which the participant did not follow instructions, i.e., performing any overt movement during the AO and MI events, or during the cue events and rest period. As a result, the percentage of excluded events was below 2 % overall, and for individual participants this percentage was always below 7 %.

The scanning session was divided into four functional runs of approximately 11 min each, with an anatomical scan interspersed after the first two functional runs and short pauses between the other runs. As shown in Figure 1, three types of trials were used during scanning: pure Action Observation (AO: video presentation followed by rest), Motor Imagery (MI: video presentation followed by motor imagery), and Action Execution (EXE: video presentation followed by imitative execution). This layout allowed us to study action observation directly followed by motor imagery or execution, whilst the pure AO condition served to minimise potential contaminations of the AO regressor by the subsequent MI or EXE events. Participants were only cued whether to rest or to engage in motor imagery or execution of the observed sequence or rhythm after the video presentation. This assured that they attentively observed each video clip regardless of condition.

In each run, 36 trials were presented consisting of 18 SEQ trials (three non-practised and three practised AO trials, three non-practised and three practised MI trials, three non-practised and three practised EXE trials) and of 18 equivalent RHY trials. Accordingly, each of the three practised spatial sequences and of the three practised rhythms was shown three times per

run, once each in an AO, MI, and EXE trial. In order to minimise opportunities for practice of the non-practised stimuli within the scanning session, the remaining sets of nine SEQ and nine RHY stimuli were used as non-practised patterns. All conditions were presented in pseudo-randomized order (for further details of the trial structure see the legend of Figure 1).

< please enter Figure 1 about here >

Data acquisition

Functional imaging was performed at 3 T MAGNETOM Trio whole-body magnetic resonance imaging scanner (Siemens Medical Systems, Erlangen, Germany) equipped with an eight-channel head coil. Thirty-two axial slices (field of view = 192 mm, 64 x 64 pixel matrix, slice thickness = 3 mm, inter-slice gap = 1.2 mm, in-plane resolution = 3 x 3 x 4.2 mm, bandwidth = 2604 Hz/Px, echo spacing = 0.45 ms) covering the whole brain from the cerebellum through to the vertex were acquired using a fast single-shot gradient echo-planar imaging (EPI)-sequence (repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°) sensitive to blood oxygenation level-dependent (BOLD) contrast. The field of view was tilted to encompass the whole brain and to avoid sinus-induced susceptibility artefacts in the frontal cortex. Four functional runs with n=333 T2*-weighted scans were performed with each scan sampling over the 32 slices. For the anatomical T1-weighted images we used a field of view = 224 mm, 224 x 256 pixel matrix, 176 slices, slice thickness = 1 mm, no inter-slice gap, in-plane resolution = 1 x 1 x 1 mm, repetition time = 2040 ms, echo time = 5.57 ms, flip angle = 8°, with SENSE factor in Parallel Acquisition Technique = 2. The total scanning time for each participant was approx. one hour.

Data analysis

Functional imaging data were analyzed using Statistical Parametric Mapping software SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/) running under Matlab 7.10 (MathWorks, Inc.; Natick, MA; USA). The first five volumes of each participant's scan were discarded to allow for T1 equilibration effects. For each participant, spatial preprocessing included realignment to the first scan, and co-registration to the T1 anatomical volume images. T1-weighted images were segmented into gray and white matter. This segmentation was the basis for spatial normalization to the Montreal Neurological Institute (MNI) template, which was then resliced and smoothed with a 9 × 9 × 9 mm full width at half maximum Gaussian Kernel filter to improve the signal-to-noise ratio. To correct for low-frequency components, a temporal high-pass filter with a cut-off frequency of 1/128 Hz (= 128 s) was applied.

Statistical analyses were performed using the general linear model as implemented in SPM8. In the first-level analysis, for each participant onsets of the action observation events across the three trial types and onsets of the motor imagery and execution events with a duration of 4.7 s were used as regressors to the model including the following 12 conditions: (1) non-practised SEQ–AO, (2) practised SEQ–AO, (3) non-practised SEQ–MI, (4) practised SEQ–MI, (5) non-practised SEQ–EXE, (6) practised SEQ–EXE, (7) non-practised RHY–AO, (8) practised RHY–AO, (9) non-practised RHY–MI, (10) practised RHY–MI, (11) non-practised RHY–EXE, (12) practised RHY–EXE. The second-level analysis was carried out using the flexible factorial design with the first two-level factor SUBJECT (non-musicians, musicians) and the second 12-level factor CONDITION (see above). For basic contrasts and conjunction analyses the significance level was set to $p < .05$, FWE-corrected for the whole brain volume. A cluster size of ≥ 20 contiguous voxels (160 mm^3) extended the threshold.

Direct contrast analyses used an uncorrected threshold of $p < .001$ with an extent of $k = 70$ voxels (560 mm^3). In order to exclude false positive activations, direct contrasts were inclusively masked by the relevant minuend contrast, thresholded at $p = 0.05$. The SPM Anatomy toolbox v1.8 (Eickhoff et al. 2005, 2007) was employed for anatomical assignments by reference to probabilistic cytoarchitectonic maps.

Results

Behavioural data

We analysed the imitation performance in the execution trials by means of a sliding window over three consecutive responses ('triplets'), starting with responses 1 to 3, then 2 to 4, etc. up to 6 to 8 (Werheid et al. 2003). The performance of any three responses in an order entailed in the correct sequence counted as one correct triplet. A correct imitation of the eight required positions (SEQ) or intervals (RHY) resulted in six correct triplets. Prior to this analysis, the raw interval durations from the rhythm trials were categorised into long, medium, and short classes using the default k-means clustering algorithm as implemented in Matlab.

Figure 2 shows the imitation performance separately for sequences and rhythms, non-practised and practised patterns, and the two groups. In the non-musicians, the non-practised sequences and rhythms were of similar difficulty, and these participants showed comparable improvements for both pattern types. The musicians showed comparable performance to the non-musicians in the sequences, whilst their imitation performance for the rhythms was substantially better. These trends were confirmed via a three-factorial ANOVA, where the main effects of task (SEQ *versus* RHY), practice, and group were highly significant, Fs (1, 26) > 22.6, ps < .001. The interactions between task and practice, task and group, and the three-

way interaction were also highly significant, *F*s (1, 26) > 18.4, *p*s < 0.001. Planned comparisons (Rosenthal and Rosnow 1985), run separately for the sequences and rhythms, indicated that the effect of practice was highly significant for each task, *F*s (1, 26) > 75.8, *p*s < 0.001. For the sequences, the effects of group and the interaction between practice and group were not significant, whilst for the rhythms both effects were highly significant, *F*s (1, 26) > 18.7, *p*s < .001. In addition, for the musicians the effect of task and the interaction between task and practice were highly significant, *F*s (1, 14) > 57.9, *p*s < .001, whilst for the non-musicians both effects were, reassuringly, non-significant. This pattern of results confirms that the musicians were selectively advantaged for rhythm imitation. In summary, the behavioural data met all prerequisites for the interpretation of the functional imaging data.

We also analysed the behavioural data separately for each triplet (n = 6) and scanning session (n = 4). As shown in Supplementary Figure 1, in the non-practised trials the first two triplets (i.e., the first four responses) were imitated with higher accuracy than the subsequent responses, indicating a primacy effect. For the practised trials, performance was clearly improved and level across the eight required positions and intervals. Importantly, these results were stable across the four sessions, as indicated by the absence of main effects of session (Fs < 1.3, ps > .30) in the related four-factorial ANOVAs (for details, see legend of Supplementary Figure 1).

< please enter Figure 2 about here >

FMRI results (1): Task networks for sequence and rhythm imitation

For the present purposes, we pragmatically define a task network as those brain regions which are jointly activated during action observation (AO) and motor execution (EXE) events.

Figure 3 and Table 1 show the related conjunction analyses separately for the SEQ and RHY tasks, each collapsed across practised and non-practised performances, and irrespective of musical expertise.

Observation and execution of the *sequences* jointly involved two extensive bilateral parieto-frontal activation clusters; the first comprising the superior and inferior parietal lobules (SPL and IPL, respectively), and the second comprising Area 6 with dorsal and ventral sectors of the precentral gyrus and the Supplementary Motor Area (SMA). In addition, we found two large subcortical activation clusters in the cerebellum and the thalamus, as well as activation foci in the pars triangularis of inferior frontal gyrus (IFG) bilaterally, where the right cluster extended to the middle frontal gyrus. There were also activations in the temporoparietal junction (TPJ) bilaterally and in the right middle and inferior temporal gyrus.

In comparison to the sequences, observation and execution of the *rhythms* jointly activated relatively small sectors of posterior parietal cortex (PPC), namely the IPL bilaterally. Rhythm-related activations were mainly found in bilateral ventral precentral gyrus (Area 6), in pars opercularis of IFG, in the SMA with a large cluster, and in the superior temporal gyrus / TPJ bilaterally. In addition, extensive subcortical activations involved the cerebellum and the basal ganglia bilaterally.

In summary, both sequence and rhythm tasks activated the classic mirror regions comprising inferior parietal and ventral premotor cortex extending to IFG, as well as the SMA and subcortical regions. Compared to the rhythm task, the sequence task activated considerably larger sectors of the PPC, and it also showed stronger activations in dorsal and ventral premotor cortex, as confirmed by a series of direct contrasts run separately for the AO and EXE events (see Supplementary Figure 2). In contrast, the rhythm task dominantly involved the superior temporal gyrus / TPJ, the SMA, and pars opercularis of IFG. Thus,

although the two task networks were not entirely distinct, we found clear differences regarding the dominant regions activated by each task across the AO and EXE events.

< please enter Figure 3 about here >

< please enter Table 1 about here >

FMRI results (2): Main effects of practice

Next, we analysed the main effects of practice, irrespective of musical expertise, by directly contrasting both non-practised > practised (np>pr), and practised > non-practised (pr>np) sequences and rhythms separately for the AO, MI, and EXE events (see Figure 4 and Supplementary Table 1). As expected, activations in most regions were stronger for the non-practised compared to the practised patterns, indicating neural efficiency effects.

During *action observation,* these practice effects for sequences and rhythms overlapped in the core fronto-parietal mirror regions. In addition, SPL and dorsal premotor cortex were dominantly activated during sequence observation, whilst superior temporal gyrus / TPJ, SMA, and IFG were dominantly activated during rhythm observation (for further details see legend of Figure 4). These practice effects corresponded closely to the two respective task networks as identified in the previous section.

During *motor imagery*, the practice effects for the sequences were more pronounced than those for the rhythms. These effects were found in bilateral IPL and in different frontal regions including the SMA, IFG, insula, anterior and middle cingulate cortex, as well as the middle frontal gyrus (MFG) bilaterally.

During *motor execution*, the practice effects for sequences and rhythms largely overlapped and included the SMA, precentral gyrus, IFG, as well as MFG, anterior and middle

cingulate cortex, and the insula. In summary, during both MI and execution, the reduced activations with practice were largely restricted to the frontal lobe and were more extensive for the sequences than for the rhythms.

Activation increases with practice. In addition to the dominant trend for neural efficiency effects reported above, we only found a small number of regions where activations increased with practice (see legend and right panels of Figure 4, and Supplementary Table 1, Sub-tables 7 to 12).

< please enter Figure 4 about here >

FMRI results (3): Cognitive control structures

We address the third and main research objective in two parts, first irrespective of musical expertise (this section), and subsequently with a focus on expertise-related effects in section 'FMRI results (4)'. Since cognitive control should be primarily required for the imitation of novel patterns and decrease with practice (Dayan and Cohen 2011), we base these analyses on contrasts of non-practised > practised patterns ('np>pr', e.g., Vogt et al. 2007, Higuchi et al. 2012). For the DLPFC, the related comparisons in the previous section did not show differential activations during *action observation*, whilst such effects were indeed present during both MI and EXE events. To recapitulate, during *motor imagery* bilateral MFG was activated more strongly for non-practised sequences, compared to the practised sequences, whilst for the rhythms, activation differences in MFG were absent. During *execution*, activation differences were present in MFG for both tasks. For sequence execution, these were found in MFG bilaterally; whilst during rhythm execution these were restricted to the right MFG (Figure 4 and Supplementary Table 1).

We extended the search for cognitive control structures by analysing regions that were jointly activated by the SEQ and RHY tasks. This contrast should indicate overlapping superordinate control mechanisms, e.g., for scheduling the relevant cognitive operations in the different events of each trial. In addition, this contrast should also reflect the overlapping regions of the two task networks. Figure 5 and Table 2 show the results of the conjunctions of the np>pr contrasts for each task separately for observation and execution.

During *action observation*, activation differences across both tasks were found in bilateral BA44 and adjacent PMv, the SMA, right BA45, bilateral middle temporal gyrus, and right IPL. These activations primarily indicate regions that were overlapping between the two task networks, as shown in Figure 3. During *motor imagery* (not shown in Figure 5), the corresponding conjunction yielded a single differential activation in the right IPL, which was coextensive with that for OBS. This reflected the sparse practice effects during MI of the rhythms.

In contrast, the conjunction across tasks for *execution* (Figure 5, bottom panel) indicated strong differential activations (np>pr) in a large cluster centred on the anterior midcingulate cortex (aMCC; Vogt 2009) and extending to the SMA, as well as in bilateral insula, IFG, and MFG. These results highlight the robust differential involvement of the aMCC and SMA and their likely role in performance monitoring across the two tasks. Henceforth, we refer to this activation cluster comprising the aMCC up to the SMA with the descriptive term ‘posterior medial frontal cortex’ (pMFC; see Discussion). By comparison, the activation differences in MFG were less prominent and only became apparent at the lower of the two statistical thresholds used for this contrast.

< please enter Figure 5 about here >

< please enter Table 2 about here >

FMRI results (4): Musical expertise

The behavioural data indicated that musical expertise primarily facilitated the encoding and imitation of the rhythms, whilst both groups showed similar results for the spatial sequences. Accordingly, we were particularly interested if the practice effects in prefrontal regions would also be modulated by musical expertise. For each event, we thus summarise the whole-brain results only briefly and consider the cognitive control hubs in greater detail. Practice effects were analysed separately by task and group, as well as via the interactions between group and practice. A more detailed account of the whole-brain results can be found in Supplementary Materials 1.

< please enter Figure 6 about here >

Action observation. During *SEQ observation*, the musicians showed relatively weak practice effects in the parieto-frontal task network, whilst they exhibited stronger and more extensive practice effects than the non-musicians for *RHY observation* in the related temporo-frontal task network, see Figure 6 and Supplementary Table 2. Regarding the cognitive control hubs, none of the four interaction contrasts between group and practice indicated group-specific effects for either the MFG or pMFC.

Motor imagery. During MI of the *SEQ patterns*, the overall activation differences for the musicians closely resembled those shown in Figure 4 for the combined groups, whilst the practice effects in the non-musicians were less extensive. More important in the present context, practice effects for the MFG and pMFC were present in each group individually, and

the related interactions did not indicate differences between groups in these regions, or in the task networks (see Supplementary Table 3). During MI of the *RHY patterns*, practice effects in the musicians were restricted to the right IPL as well as bilateral cerebellum, and in the non-musicians practice effects were practically absent. It is thus not surprising that differential activations in MFG and pMFC were also absent during rhythm imagery in both groups.

Execution. As expected, both groups showed similar practice effects on the whole-brain level during *SEQ execution*. Furthermore, both pMFC and bilateral MFG were differentially activated in each group individually (see white circles in Figure 6 and Supplementary Table 4). In contrast, during *RHY execution* the musicians exhibited weaker and less extensive practice effects than the non-musicians. Here, the MFG was only differentially activated in the non-musicians. This pattern of results is mirrored in the parameter estimates for MFG (Supplementary Figure 3, bottom panels), and it essentially reflects the rhythm-specific expertise of the musicians.

However, expertise-related differences during motor execution were not found for the pMFC, which was absent in the related interaction contrasts (Supplementary Table 4). Also the parameter estimates for the pMFC indicate equivalent practice effects for SEQ and RHY in both groups (Supplementary Figure 3, panels for anterior cingulate cortex and SMA). Thus, whilst the pMFC exhibited more robust practice effects in the cross-task conjunction than the MFG (Figure 5), only the MFG reflected the task-specific expertise effects observed in the behavioural data.

Discussion

This study makes three main contributions: one to the literature on mirror mechanisms, and the other two regarding the cognitive control structures involved in imitation learning. *First*, the two sequencing tasks engaged task networks which partially overlapped but which also substantially dissociated. Given that both tasks were carefully matched for difficulty and visual appearance, our data provide striking support for the concept of *task-specific mirror mechanisms* (Subiaul 2010; Rizzolatti et al. 2014). *Second*, we found that the DLPFC was involved during motor imagery of the sequences, but not for the rhythms, *thus providing fresh support for Buccino et al.'s (2004) model of imitation learning*. The DLPFC was also involved during execution of both tasks, indicating a wider, less task-specific role during motor execution. *Third*, the posterior medial frontal cortex (pMFC), known for its role in performance monitoring, was also involved during imitative execution of the SEQ and RHY tasks, where activations were more pronounced than those in DLPFC. This dominant involvement of the pMFC in the present study, compared to the dominant role of the DLPFC in the imitation learning of hand postures (e.g., Buccino et al. 2004), indicates that *the dominant cognitive control hubs for imitation learning can also vary with the task*. In addition to these three main findings, we replicated and extended earlier results regarding neural efficiency effects in action observation and execution, and regarding the effects of musical expertise on imitation performance.

Behavioural data: Effects of practice and musical expertise

The behavioural data of imitation performance in the scanner (Figure 2) provide a crucial background for the interpretation of the functional data. Results confirmed that (1) SEQ and RHY patterns were equally difficult for the non-musicians, (2) the practice effects were comparable across the two tasks, (3) the musicians were only marginally more accurate than

the non-musicians in sequence imitation, and (4) the musicians were substantially more accurate than the non-musicians in the imitation of novel and practised rhythms, confirming the domain-specificity of expertise (Chase and Simon 1973; see also Matthews et al. 2016).

Further analysis of the behavioural data confirmed that no substantial learning occurred within the scanning session. This likely resulted from the randomised order of patterns across trials during scanning, and from the use of a sufficiently large pool of non-practised patterns. Finally, we found that participants' imitation accuracy was initially not uniform across the eight positions or intervals. Instead, for the non-practised patterns, the first four responses were performed with greater accuracy than the subsequent ones, whereas accuracy was consistently high across all responses for the practised patterns. Most likely, participants had learned to group the observed elements of a given sequence, as well as their responses, into larger units or 'chunks' (Gobet et al. 2001; Keele et al. 2003; Hard et al. 2011).

Dissociable task networks for sequence and rhythm imitation

We begin the discussion of the imaging data with the two *task networks* (*research objective 1*), defined here as the activated areas during both observation and execution. In the two subsequent sections, we consider the effects of practice (*research objective 2a*) and expertise (*research objective 2b*) within the task networks. On this basis, we then proceed to discuss the effects of practice and expertise on the cognitive control structures (*research objectives 3a and 3b*), separately for the DLPFC and the pMFC.

Spatial sequence imitation. The task network for SEQ imitation essentially comprised the SMA, PMv and dorsal premotor cortex (PMd), large sectors of the PPC, smaller sectors in temporal cortex and in the pars triangularis of IFG, and the cerebellum (Figure 3). In particular, PMv and IPL form the classic fronto-parietal mirror circuit (Rizzolatti et al. 2014), and PMd

and SPL have been reported as a separate, reaching-related mirror circuit (DiDio et al. 2013; Filimon et al. 2015). In addition, the SMA is one of the regions for which mirror properties have been shown via single-cell recordings in the human brain (Mukamel et al. 2010), and its role in sequence learning is well-documented (Dayan and Cohen 2011). Our results are therefore consistent with the existing work on action observation and on the imitation learning of hand postures (Buccino et al. 2004; Vogt et al. 2007).

Rhythm imitation. The SEQ and RHY task networks overlapped in the PMv, IPL, SMA and cerebellum (Figure 3). Differences between tasks were observed in the pars opercularis of IFG (as part of Broca's region), the TPJ, the SMA, and the left insula, where rhythm imitation evoked stronger activations than the SEQ task (Supplementary Figure 2 and Table 1). In contrast, the SEQ task engaged the premotor regions more strongly, as well as considerably larger sectors of the PPC. In summary, whilst the SEQ task showed remarkable overlap with the posture imitation task of Buccino et al. (2004), and whilst all three tasks (SEQ, RHY, and posture imitation) exhibited overlap with respect to the fronto-parietal mirror circuit, the RHY task further recruited a different network essentially comprising Broca's region and the TPJ.

A tentative explanation for this partial dissociation between the SEQ and RHY tasks is that participants employed different components of working memory (Baddeley 2010). Encoding a sequence of locations is a classic task associated with visuo-spatial working memory. In contrast, rhythmical patterns are typically encoded in a separate, auditory working memory system for phonological, rhythmical-temporal, and pitch information (Schulze and Koelsch 2012). For example, Hickok et al. (2003) found two main regions activated for listening and covert rehearsal of both speech and rhythmical melodies, namely a region in the left posterior Sylvian fissure at the TPJ, as well as Broca's region. Both regions are coextensive with the present, RHY-specific task network. Interestingly, we found this overlap

between Hickok et al.'s and our results even though we had presented, for reasons of comparability between tasks, the RHY task in the visual modality. A plausible explanation is that our participants recoded the visual rhythms into subvocal articulatory gestures (for example, 'da, da, daaa, da-da-da-da, da-da' for M, M, L, S, S, S, M, S), which made the rhythms accessible to the auditory working memory system. Indeed, the majority of participants in either group reported that they memorised the rhythms using such covert articulations. Since Broca's region and TPJ were already involved during action observation, it is likely that participants recoded the visual gestures into subvocal articulatory gestures *on-line*, that is, whilst observing the visual rhythms.

To summarise, we suggest that the task network for rhythm imitation consists of two sensory-motor circuits, (1) the initial visuo-motor encoding of the observed finger movements in the fronto-parietal mirror circuit, from which (2) the movements are recoded on-line as subvocal articulatory gestures in an auditory working memory circuit comprising Broca's region and the TPJ (Hickok et al. 2003, see also Lahav et al. 2007). In line with Haslinger et al. (2005), who reported the recruitment of auditory areas during pianists' observation of silent piano playing, our findings can be interpreted as transmodal sensorimotor encoding (for a general framework for simultaneous processes of AO and MI, see Vogt et al. 2013). As in Haslinger et al.'s study, our musicians showed stronger practice effects in the Broca-TPJ circuit than the non-musicians. The fact that our non-musicians also engaged in this recoding is most likely due to the relatively simple visual rhythms in the present study for which musical expertise is not essential.

Whilst delineating the precise mechanisms of transmodal sensorimotor encoding of visually presented rhythms is beyond the scope of the present study, the partial dissociation of the SEQ and RHY task networks is in itself an interesting and important finding: It supports

the concept of task-specific mirror mechanisms (Subiaul 2010; Rizzolatti et al. 2014, p. 671) in a single experiment using visually well-matched action stimuli. For example, Abdollahi et al. (2013) recently reported action-specific processing in PPC for observation of climbing and object manipulation.

Activation changes with practice in the task networks

The main purpose of contrasting non-practised and practised patterns, as well as the purpose of contrasting non-musicians and musicians, was to assess the differential involvement of cognitive control structures (see dedicated discussion sections below). For this reason, we keep the discussion of practice and expertise effects *on the task networks* brief (a more detailed account can be found in Supplementary Materials 2, where we also link these findings to the literature on sequence learning).

First, across groups and AO, MI, and EXE events, most regions of the SEQ and RHY task networks exhibited neural efficiency effects, that is, stronger activations for the non-practised patterns than for the practised patterns (Figure 4). In contrast, increases with practice were sparse, and the ratio of activated voxels showing neural efficiency effects, relative to those exhibiting increases with practice, exceeded 4:1 in all comparisons displayed in Figure 4 and Supplementary Table 1. A similar prevalence of practice-related activation decreases was reported by Vogt et al. (2007) and Higuchi et al. (2012), where the literature on practice effects during action observation is discussed in greater detail. *Second*, the neural efficiency effects for each task essentially mirrored the two task networks as identified in the previous section (compare Fig. 3 and 4 and related Tables). This provides convergent evidence for the partial dissociation of the SEQ and RHY task networks. *Third*, during both MI and EXE events, the neural efficiency effects were predominantly found in the frontal lobe. Again they

resembled the related sectors of the two task networks, and they were more extensive for the sequences than for the rhythms. Overall, these practice effects are consistent with the available literature on 'fast' sequence learning (for details, see Supplementary Materials 2). Importantly, also the MFG and pMFC showed significantly reduced activations with practice during MI and EXE events (see discussion of cognitive control structures). *Fourth,* the practice effects for MI clearly dissociated from those during AO and were a fair subset of those during execution (Figure 4). This activation overlap between MI and EXE is in line with the widely accepted view of motor imagery as a form of motor simulation that engages neural structures used in execution (Jeannerod 2001; Vogt et al. 2013). In the interest of brevity, we reserve an in-depth comparison of the activation differences between AO, MI, and EXE for a separate report.

Expertise-related practice effects in the task networks

Action observation. As shown in Figure 6, the results for the *non-musicians* largely resembled the results across groups (Figure 4) for both tasks. One difference was that during rhythm observation the practice effects for the Broca-TPJ circuit were less extensive, although clearly present. In contrast, the *musicians* exhibited more extensive neural efficiency effects during rhythm observation, whilst they exhibited considerably less extensive effects than the non-musicians during sequence observation. The stronger activations for rhythm observation in the musicians, both in direct comparison to the non-musicians for novel rhythms and when comparing the neural efficiency effects between groups, replicate expertise effects as demonstrated in earlier studies (e.g., Haslinger et al. 2005; Calvo-Merino et al. 2005, 2006). In addition, the present study highlights a clear functional role of the musicians' enhanced activations during rhythm observation, namely to enable their exquisite imitation performance

in subsequent execution. As such, the present results demonstrate *experts' enhanced capacity to encode novel observed actions for subsequent imitation in their domain of expertise.*

Motor imagery. In the task networks, the musicians tended to show more extensive activation differences during MI than the non-musicians. Apart from this trend, the group differences during MI were negligible.

Execution. Again, both participant groups showed similar results for spatial sequence execution. In contrast, for the rhythms the musicians showed less extensive neural efficiency effects than the non-musicians (Supplementary Table 4 and Figure 6, bottom panels) in the cerebellum, sensorimotor cortex, right superior and middle frontal gyrus, angular gyrus, and insula.

In summary, compared to the non-musicians, the musicians exhibited particularly strong activations during observation of the novel rhythms, associated with more extensive practice effects in the related task network. This set of findings is in line with earlier research on expertise effects in action observation (e.g., Haslinger et al. 2005; Calvo-Merino et al. 2005, 2006), In addition, it highlights experts' enhanced *capacity* for visuo-motor encoding during action observation in the context of imitation. During subsequent execution, the musicians showed relatively small differences between non-practised and practised rhythms, which we would interpret as a 'pay-off' related to the enhanced processing during rhythm observation. We shall revisit this rhythm-specific asymmetry between groups in the context of cognitive control structures, to which we turn next.

Dorsolateral prefrontal cortex in motor imagery and execution

The main motivation for the present study was to explore the involvement of the DLPFC and other cognitive control structures in the imitation learning of spatial sequences and rhythms

(*research objective 3*). Since cognitive control is primarily required in the early stages of learning and reduces with practice (Kelly and Garavan 2005; Dayan and Cohen 2011), we assessed this via the within-session activation differences between non-practised and practised patterns (see also Vogt et al. 2007; Higuchi et al. 2012). The analyses of practice effects across groups, both conjunct and run separately for each task, consistently revealed no differential activations during action observation. During motor imagery, practice effects were found for the SEQ task but not for the RHY task, and during execution, practice effects were present in DLPFC bilaterally for the SEQ task and in right DLPFC for rhythm execution. When the practice effects were examined separately for each group, during action observation DLPFC was found differentially activated only in a small cluster when the musicians observed the rhythms. During motor imagery, again each group showed differential practice effects for the sequences only. During execution, activations in DLPFC reduced bilaterally with practice in each group for the sequences, whilst during rhythm execution only the non-musicians showed this effect reliably, where it was largely right-lateralised (see also parameter estimates in Supplementary Figure 3, bottom panels). These results inform Buccino et al.'s (2004) model of imitation learning in the following ways:

First, the paucity of DLPFC activations during *action observation* is not entirely surprising: in the present SEQ and RHY tasks, action observation primarily required the sustained encoding of the sequence of stimuli throughout the observation interval, which provided little opportunity for cognitive control. In contrast, in the posture imitation studies by Buccino et al. (2004) and Vogt et al. (2007), participants watched the same hand posture over a period of 4 to 10 s, which allowed them to apply various cognitive-exploratory strategies already during action observation, as well as during the subsequent motor preparatory period.

This was reflected in the differential practice effects in DLPFC previously found for these two events (Vogt et al. 2007).

Second, DLPFC was differentially activated during *motor imagery* of the sequences, but not for the rhythms. This second main finding of the present study provides an important extension of Buccino et al.'s (2004) two-level model of imitation learning, namely to spatial sequences. A number of qualifications are appropriate here. In a given trial, our participants either engaged in MI or in imitative execution, but not in both in direct succession (see Figure 1). We had chosen this design in order to eliminate possible contaminations of the BOLD signal between the two events. In contrast, Buccino et al. (2004) and Vogt et al. (2007) inserted a motor preparatory event between observation and execution. Whilst it is likely that participants engaged in MI in both situations, this cannot be known for certain for the two earlier studies. In addition, further behavioural research will be required to establish to what extent such a preparatory / MI period actually facilitates imitation learning behaviourally. In the present study, participants were certainly capable of imitating immediately after action observation (see also Vogt 1996; Higuchi et al. 2012), however, the absence of between-session effects for the non-practised patterns in the behavioural data might indicate that such a "see – do" scenario is not particularly suitable for supporting learning. For the time being, we would thus maintain that a preparatory / MI interval facilitates imitation learning, by allowing for the mental rehearsal and cognitive control of the to-be executed action. The present study then suggests the involvement of DLPFC as a likely neural mechanism. Its primary role is most likely not the maintenance of visuo-spatial information but rather the selection and preparation of such information for motor execution (Pochon et al. 2001; Passingham and Sakai 2004; Sakai 2008), as well as potentially the monitoring of MI (see below).

Third, DLPFC was not activated during MI of the rhythms. Interestingly, in their elegant TMS study, Robertson et al. (2001) found that the critical role of the DLPFC in their sequence learning task was also restricted to spatially cued sequences. Taken together, these findings indicate a possible qualification of Buccino et al.'s (2004) model of imitation learning, which was solely based on the imitation of hand postures: According to the available evidence, the supervisory role of DLPFC during motor preparation (Buccino et al. 2004; Vogt et al. 2007) and motor imagery (this study) is likely restricted to visuo-spatial patterns. Indeed, whilst in principle, a sequence of locations can be cognitively manipulated (e.g., interrupted, corrected and 'restarted'), such operations are more difficult to apply to rhythmical patterns, as they are defined by their temporal structure. This might also explain the relatively small overall practice effects during MI of the rhythms. The dissociation between spatial and rhythmical patterns, as reported here regarding prefrontal involvement, also informs future meta-analytic work. For example, in the meta-analysis of MI by Hétu et al. (2013), MFG was found to be involved during MI of motor sequences, but no distinction between spatial and rhythmical sequences was made.

Fourth, the involvement of DLPFC during *execution* of the present SEQ and RHY tasks presumably reflects sustained monitoring and cognitive control throughout imitative execution. Shallice (2004) proposed that the right DLPFC is primarily involved in monitoring whether a newly configured motor plan is executed in accordance with the task goals. The right-hemispheric dominance of the present DLPFC activations suggests that DLPFC was indeed primarily engaged in monitoring motor execution (see also Vogt et al. 2007).

Finally, the execution-related practice effects in DLPFC were similarly pronounced in both groups for the SEQ task, but for the RHY task, they were reduced in the musicians, compared to the non-musicians (Figure 6). These results mirror the behavioural findings,

where the musicians were selectively advantaged in imitating particularly the non-practised rhythms (Figure 2), and they further resemble the pattern of activation differences in the task networks. Whilst it might seem straightforward to attribute the null results for the DLPFC to the musicians' expertise in rhythm processing (Matthews et al. 2016), the activations during the immediately preceding action observation event require a qualification of this interpretation: As discussed in the previous section, during AO the musicians exhibited particularly strong differential activations in the rhythm task network, as well as in a small sector of the DLPFC. It is therefore also viable to interpret the musicians' reduced practice effects during rhythm execution, in both the task network and DLPFC, as a 'pay-off' of the strong differential activations in this group during rhythm observation.

Posterior medial frontal cortex and performance monitoring

Apart from the DLPFC, the pMFC is the other major cognitive control hub that was found activated in the present study. With the descriptive term pMFC, we refer primarily to the core regions aMCC (Vogt 2009) and pre-SMA, as well as adjacent SMA, which have been found co-activated in many neuroimaging experiments (Ridderinkhof et al. 2004; Ullsperger et al. 2014). During AO, we found practice-related activation differences in the SMA but not in cingulate cortex (this was confirmed by the conjunction analyses in Figure 3 and Table 1). During MI of the spatial sequences, activations included not only the DLPFC but also the pMFC (i.e. aMCC and SMA regions), and during motor execution, pMFC was saliently differentially activated for both SEQ and RHY tasks. We regard the robust involvement of the pMFC during motor execution of both tasks as the third main finding of the present study.

First of all, the possible functions of the pMFC in cognitive control have been extensively studied over the last two decades using a variety of electrophysiological and brain

imaging techniques (Ullsperger and von Cramon 2004; Ridderinkhof et al. 2004; Ullsperger et al. 2014), where experimental paradigms were typically designed to probe, e.g., error detection *versus* conflict monitoring, independently of motor skill learning. Whilst the precise functions of pMFC are still under debate, its general role as a major cognitive control structure involved in performance monitoring is now widely accepted. In the context of skill learning, the anterior cingulate cortex, along with lateral prefrontal and posterior parietal cortices, is generally considered to perform a scaffolding role (Kelly and Garavan 2005). Indeed, the transient involvement of the cingulate cortex, along with the DLPFC, in the early stages of sequence learning was recently demonstrated by Basset et al. (2015, see Introduction).

In the present study, the activations in pMFC can be very well interpreted *sensu* performance monitoring. During *action observation*, participants primarily engaged in sustained encoding of the stimuli, and no activations of cingulate cortex were found during this event, consistent with previous neuroimaging studies (see Buccino et al. 2004; Caspers et al. 2010). We have already interpreted the engagement of the SMA (proper) during AO as part of the task network related to sequence encoding.

The sustained activation of the task networks (including the fronto-parietal mirror circuit) across AO and EXE stands in contrast to the exclusive engagement of the pMFC during *MI and execution*. In the present tasks, performance monitoring likely included a number of processes. First, in the practice session most, if not all participants had detected the common features across all sequences and rhythms used. These included the fixed number of positions and intervals (n = 8), as well as certain regularities, such as no repetition of positions within the first four and the last four SEQ elements. In the scanner, participants could then check their performances (physical or imagined) against these general features. Second, they might have occasionally detected a mismatch between their sensorimotor representation of a

just-observed pattern and their execution. Third, the generation of the relatively long patterns might involve a more general requirement for sustained performance monitoring throughout MI and execution, independent of error monitoring.

The practice-related activation differences in pMFC during motor execution were more robust than those in the DLPFC. In the related cross-task conjunction (Figure 5), only the pMFC activations, along with left Broca's region and the insula, passed the more conservative of the two statistical thresholds. In addition, the related parameter estimates (Supplementary Figure 3) were generally higher for the cingulate cortex and the SMA region than for the DLPFC. This result indicates that not only the task networks can vary according to task demands, but also that the dominant cognitive control structures can vary. In contrast to the imitation of hand postures (e.g., Buccino et al. 2004), the sequential tasks used in the present study presumably render themselves more readily for performance monitoring than for restructuring operations in both motor imagery and execution. In fact, we have already interpreted the right-dominant involvement of the DLPFC to reflect monitoring operations, rather than primarily restructuring (Shallice, 2004). Alternatively, Ridderinkhof et al. (2004, p. 443) proposed a possible division of labour between pMFC and the DLPFC, namely that "monitoring-related pMFC activity serves as a signal that engages regulatory processes in the lateral prefrontal cortex to implement performance adjustments". Although we have no direct evidence that this would apply to the present study, this is certainly an attractive working hypothesis.

Conclusions

The present research provides an important extension to earlier studies on imitation learning

(Buccino et al. 2004; Vogt et al. 2007; Higuchi et al. 2012). Whilst we found that the fronto-parietal mirror circuit was involved in both SEQ and RHY tasks, sequence imitation relied more strongly on posterior parietal regions, and rhythm imitation recruited an additional task network for encoding rhythmical-temporal information (Schulze and Koelsch 2012). This partial dissociation supports the concept of task-specific mirror mechanisms (Subiaul 2010; Rizzolatti et al. 2014). We were also able to further specify the involvement of cognitive control structures. During motor imagery, the DLPFC showed practice-related modulations for the SEQ task, thus extending Buccino et al.'s (2004) two-level model spatial sequences. In contrast, no such practice effects were found during motor imagery of the rhythms. Both pMFC and DLPFC were strongly involved during the imitative execution of spatial sequences and rhythms. Both regions are well-known as cognitive control hubs, and the present results suggest a dominant role of the pMFC, commensurate with its crucial role of performance monitoring in sequence execution. Finally, the musicians exhibited an enhanced capacity for encoding the novel rhythms during AO, which payed-off in their exquisite subsequent imitation performance.

In their initial study on the topic, Buccino et al. (2004, p. 331) concluded that their 'minimalistic' interpretation of the anatomical basis of imitation learning "does not exclude that in imitation conditions where other aspects of the action to be imitated (such as a sequence or rhythm) are fundamental, a crucial role is played also by neural structures other than those evidenced in the present study". Indeed, the present results testify that the neural mechanisms of imitation learning reflect first and foremost (a) the anatomical structures involved in the specific motor task under study, and (b) the task-relevant cognitive control structures. In particular, the robust involvement of the pMFC in the present study nicely corrobates Heyes' (2009, p. 2295) proposal that "imitation learning enlists additional, general

purpose mechanisms of learning and cognitive control" rather than mechanisms restricted to imitation. A task for future research will be to characterise the nature of the interactions between different cognitive control structures, and between these and specific task networks, in imitation learning.

Funding

This work was supported by a research project grant from the Leverhulme Trust (grant number F/00 185/K) to S.V. and N.R. The authors declare no competing financial interests.

Acknowledgements

We would like to thank Giacomo Rizzolatti (Parma, Italy) for many fruitful discussions which inspired the present study, and Valerie Adams, Anna Anderson, and Bill Bimson (Liverpool), Dave Gaskell and Gordon Johnson (Lancaster) for their kind technical and administrative support during setup and running of this study. Virginia Kellond (Liverpool) helped with participant-management and with scoring the video recordings, and Ryssa Moffat (Ottawa, Canada) proofread the manuscript.

References

Abdollahi RO, Jastorff J, Orban GA. 2013. Common and segregated processing of observed actions in human SPL. Cereb Cortex. 23(11):2734–2753.

Babiloni C, Del Percio C, Rossini PM, Marzano N, Iacoboni M, Infarinato F, Lizio R, Piazza M, Pirritano M, Berlutti G, Cibelli G, Eusebi F. 2009. Judgment of actions in experts: a high-resolution EEG study in elite athletes. Neuroimage. 45(2):512–521.

Babiloni C, Marzano N, Infarinato F, Iacoboni M, Rizza G, Aschieri P, Cibelli G, Soricelli A, Eusebi F, Del Percio C. 2010. "Neural efficiency" of experts' brain during judgment of actions: a high-resolution EEG study in elite and amateur karate athletes. Behav Brain Res. 207(2):466–475.

Baddeley A. 2010. Working memory. Curr Biol. 20(4):R136–140.

Bassett DS, Yang M, Wymbs NF, Grafton ST. 2015. Learning-induced autonomy of sensorimotor systems. Nat Neurosci. 18(5):744–751.

Buccino G, Vogt S, Ritzl A, Fink GR, Zilles K, Freund HJ, Rizzolatti G. 2004. Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. Neuron. 42(2):323–334.

Byrne RW. 2003. Imitation as behaviour parsing. Philos Trans R Soc Lond B Biol Sci. 358(1431):529–536.

Calvo-Merino B, Glaser DE, Grèzes J, Passingham RE, Haggard P. 2005. Action observation and acquired motor skills: an FMRI study with expert dancers. Cereb Cortex. 15(8):1243–1249.

Calvo-Merino B, Grèzes J, Glaser DE, Passingham RE, Haggard P. 2006. Seeing or doing? Influence of visual and motor familiarity in action observation. Curr Biol. 16(19):1905–1910. Erratum in: Curr Biol. 16(22):2277.

Caspers S, Zilles K, Laird AR, Eickhoff SB. 2010. ALE meta-analysis of action observation and imitation in the human brain. Neuroimage. 50(3):1148–1167.

Chartrand TL, van Baaren R. 2009. Human mimicry. In: Zanna MP, editor. Advances in experimental social psychology. Volume 41. Amsterdam (NL); London/Oxford (UK); Burlington (MA); San Diego (CA): Academic Press/Elsevier. p 219–274.

Chase WG, Simon HA. 1973. Perception in chess. Cognit Psychol. 4:55–81.

Dayan E, Cohen LG. 2011. Neuroplasticity subserving motor skill learning. Neuron. 72(3):443–454.

Di Dio C, Di Cesare G, Higuchi S, Roberts N, Vogt S, Rizzolatti G. 2013. The neural correlates of velocity processing during the observation of a biological effector in the parietal and premotor cortex. Neuroimage. 64:425–436.

Doyon J, Benali H. 2005. Reorganization and plasticity in the adult brain during learning of motor skills. Curr Opin Neurobiol. 15(2):161–167.

Eickhoff SB, Paus T, Caspers S, Grosbras MH, Evans A, Zilles K, Amunts K. 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage. 36(3):511–521.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage. 25(4):1325–1335.

Filimon F, Rieth CA, Sereno MI, Cottrell GW. 2015. Observed, executed, and imagined action representations can be decoded from ventral and dorsal areas. Cereb Cortex. 25(9):3144–3158.

Gobet F, Lane PC, Croker S, Cheng PC, Jones G, Oliver I, Pine JM. 2001. Chunking mechanisms in human learning. Trends Cogn Sci. 5(6):236–243.

Halsband U, Lange RK. 2006. Motor learning in man: a review of functional and clinical studies. J Physiol Paris. 99(4-6):414–424.

Hamilton, AF. 2015. The neurocognitive mechanisms of imitation. Curr Opin Behav Sci. 3:63–67.

Hard BM, Recchia G, Tversky B. 2011. The shape of action. J Exp Psychol Gen. 140(4):586–604.

Hardwick RM, Rottschy C, Miall RC, Eickhoff SB. 2013. A quantitative meta-analysis and review of motor learning in the human brain. Neuroimage. 67:283–297.

Haslinger B, Erhard P, Altenmüller E, Schroeder U, Boecker H, Ceballos-Baumann AO. 2005. Transmodal sensorimotor networks during action observation in professional pianists. J Cogn Neurosci. 17(2):282–293.

Hétu S, Grégoire M, Saimpont A, Coll MP, Eugène F, Michon PE, Jackson PL. 2013. The neural network of motor imagery: an ALE meta-analysis. Neurosci Biobehav Rev. 37(5):930–949.

Heyes C. 2009. Evolution, development and intentional control of imitation. Philos Trans R Soc Lond B Biol Sci. 364(1528):2293–2298.

Hickok G, Buchsbaum B, Humphries C, Muftuler T. 2003. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci. 15(5):673–682.

Higuchi S, Holle H, Roberts N, Eickhoff SB, Vogt S. 2012. Imitation and observational learning of hand actions: prefrontal involvement and connectivity. Neuroimage. 59(2):1668–1683.

Jeannerod M. 2001. Neural simulation of action: a unifying mechanism for motor cognition. Neuroimage. 14(1 Pt 2):S103–109.

Keele SW, Ivry R, Mayr U, Hazeltine E, Heuer H. 2003. The cognitive and neural architecture of sequence representation. Psychol Rev. 110(2):316–339.

Kelly AM, Garavan H. 2005. Human functional neuroimaging of brain changes associated with practice. Cereb Cortex. 15(8):1089–1102.

Kilner JM, Friston KJ, Frith CD. 2007. Predictive coding: an account of the mirror neuron system. Cogn Process. 8(3):159–166.

Kilner JM, Lemon RN. 2013. What we know currently about mirror neurons. Curr Biol. 23(23):R1057–1062.

Konoike N, Kotozaki Y, Miyachi S, Miyauchi CM, Yomogida Y, Akimoto Y, Kuraoka K, Sugiura M, Kawashima R, Nakamura K. 2012. Rhythm information represented in the fronto-parieto-cerebellar motor system. Neuroimage. 63(1):328–338.

Koski L, Iacoboni M, Dubeau MC, Woods RP, Mazziotta JC. 2003. Modulation of cortical activity during different imitative behaviors. J Neurophysiol. 89(1):460–471.

Lahav A, Saltzman E, Schlaug G. 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci. 27(2):308–314.

Lohse KR, Wadden K, Boyd LA, Hodges NJ. 2014. Motor skill acquisition across short and long time scales: a meta-analysis of neuroimaging data. Neuropsychologia. 59:130–141.

Matthews TE, Thibodeau JN, Gunther BP, Penhune VB. 2016. The impact of instrument-specific musical training on rhythm perception and production. Front Psychol. 7:69.

Molenberghs P, Cunnington R, Mattingley JB. 2012. Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. Neurosci Biobehav Rev. 36(1):341–349.

Mukamel R, Ekstrom AD, Kaplan J, Iacoboni M, Fried I. 2010. Single-neuron responses in humans during execution and observation of actions. Curr Biol. 20(8):750–756.

Nissen MJ, Bullemer P. 1987. Attentional requirements of learning: evidence from performance measures. Cognit Psychol. 19(1):1–32.

Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9(1):97–113.

Passingham D, Sakai K. 2004. The prefrontal cortex and working memory: physiology and brain imaging. Curr Opin Neurobiol. 14(2):163–168.

Petersen SE, van Mier H, Fiez JA, Raichle ME. 1998. The effects of practice on the functional anatomy of task performance. Proc Natl Acad Sci U S A. 95(3):853–860.

Pochon JB, Levy R, Poline JB, Crozier S, Lehéricy S, Pillon B, Deweer B, Le Bihan D, Dubois B. 2001. The role of dorsolateral prefrontal cortex in the preparation of forthcoming actions: an fMRI study. Cereb Cortex. 11(3):260–266.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. 2004. The role of the medial frontal cortex in cognitive control. Science. 306(5695):443–447.

Rizzolatti G. 2014. Imitation: mechanisms and importance for human culture. Rend Fis Acc Lincei. 25(3):285–289.

Rizzolatti G, Cattaneo L, Fabbri-Destro M, Rozzi S. 2014. Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. Physiol Rev. 94(2):655–706.

Robertson EM, Tormos JM, Maeda F, Pascual-Leone A. 2001. The role of the dorsolateral prefrontal cortex during sequence learning is specific for spatial information. Cereb Cortex. 11(7):628–635.

Rosenthal R, Rosnow RL. 1985. Contrast analysis: focused comparisons in the analysis of variance. Cambridge (UK): Cambridge University Press.

Sakai K. 2008. Task set and prefrontal cortex. Annu Rev Neurosci. 31:219–245.

Schulze K, Koelsch S. 2012. Working memory for speech and music. Ann N Y Acad Sci. 1252:229–236.

Shallice T. 2004. The fractionation of supervisory control. In: Gazzaniga MS, editor. The cognitive neurosciences. 3rd ed. Cambridge (MA): MIT Press. p 943–956.

Subiaul F. 2010. Dissecting the imitation faculty: the multiple imitation mechanisms (MIM) hypothesis. Behav Processes. 83(2):222–234.

Ullsperger M, Danielmeier C, Jocham G. 2014. Neurophysiology of performance monitoring and adaptive behavior. Physiol Rev. 94(1):35–79.

Ullsperger M, von Cramon DY. 2004. Neuroimaging of performance monitoring: error detection and beyond. Cortex. 40(4–5):593–604.

Vogt BA. 2009. Regions and subregions of the cingulate cortex. In: Vogt BA, editor. Cingulate neurobiology and disease. Oxford (UK): Oxford University Press. p 3-30.

Vogt S. 1996. Imagery and perception-action mediation in imitative actions. Brain Res Cogn Brain Res. 3(2):79–86.

Vogt S. 2002. Visuomotor couplings in object-oriented and imitative actions. In: Meltzoff AN, Prinz W, editors. The imitative mind: development, evolution, and brain bases. Cambridge (UK): Cambridge University Press. p 206–220.

Vogt S, Buccino G, Wohlschläger AM, Canessa N, Shah NJ, Zilles K, Eickhoff SB, Freund HJ, Rizzolatti G, Fink GR. 2007. Prefrontal involvement in imitation learning of hand actions: effects of practice and expertise. Neuroimage. 37(4):1371–1383.

Vogt S, Di Rienzo F, Collet C, Collins A, Guillot A. 2013. Multiple roles of motor imagery during action observation. Front Hum Neurosci. 7:807.

Vogt S, Thomaschke R. 2007. From visuo-motor interactions to imitation learning: behavioural and brain imaging studies. J Sports Sci. 25(5):497–517.

Wang Y, Hamilton AF. 2012. Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. Front Hum Neurosci. 6:153.

Werheid K, Ziessler M, Nattkemper D, von Cramon DY. 2003. Sequence learning in Parkinson's disease: the effect of spatial stimulus-response compatibility. Brain Cogn. 52(2):239–249.

Tables

Table 1. Task networks for sequence and rhythm imitation.

Table 2. Conjunctions between sequence and rhythm tasks.

Table 1. Task networks for sequence and rhythm imitation. Macroanatomical structure, cytoarchitectonical area ($Area_{cyto}$), percent overlap of cluster with cytoarchitectonical area, cluster size in voxel, MNI coordinates (x, y, z), and maximum T value (T_{max}) of the local maxima of the conjunctions between action observation (AO) and execution (EXE), separately for spatial sequences (SEQ) and rhythms (RHY). Analyses included both groups, and non-practised and practised patterns. The significance level was set to $p < .05$, FWE-corrected. A cluster size of ≥ 20 contiguous voxels (160 mm^3) extended the threshold. Abbreviations: L. = left, R. = right, TPJ = temporoparietal junction.

				MNI coordinates			
Local maximum in macroanatomical structure	**$Area_{cyto}$**	**Percent overlap of cluster with cytoarchitectonical area**	**Cluster size (voxel)**	**x**	**y**	**z**	**T_{max}**
(1) SEQ: AO ∩ EXE (non-practised + practised)							
R. Superior Parietal Lobule	SPL (7A)	6.2	9025	20	-56	60	20.18
R. Inferior Parietal Lobule*	Area 2	7.2		36	-42	46	19.23
L. Inferior Parietal Lobule*	hIP3	3.0		-38	-38	42	19.16
L. Superior Parietal Lobule*	SPL (7A)	10.0		-24	-54	60	18.83
L. Superior Parietal Lobule*	Area 2	6.1		-34	-48	56	17.64
R. Superior Parietal Lobule*	Area 2	7.2		32	-48	56	16.30
L. Precentral Gyrus	Area 6	17.9	7867	-28	-8	54	21.87
R. Precentral Gyrus*				26	-6	52	20.50

L. Precentral Gyrus*				-52	2	30	16.81
R. Supplementary Motor Area (SMA)*	Area 6	11.2		8	8	46	13.03
R. Precentral Gyrus*				52	6	32	11.66
L. Supplementary Motor Area (SMA)*	Area 6	17.9		-6	-2	58	11.29
R. Cerebellum	Lobule VI	18.1	4741	34	-58	-26	13.16
L. Cerebellum*	Lobule VI	18.4		-30	-62	-26	11.51
L. Thalamus	Th-Prefrontal	11.0	4031	-10	-22	8	12.34
R. Thalamus*	Th-Prefrontal	8.6		10	-18	8	10.59
L. Inferior Frontal Gyrus (Pars Triangularis)			169	-40	26	24	7.73
R. Superior Temporal Gyrus / TPJ	IPC (PF)	78.4	95	60	-36	18	8.38
R. Inferior Frontal Gyrus (Pars Triangularis)			75	44	28	26	6.25
R. Middle Frontal Gyrus*				46	32	22	5.75
R. Middle Temporal Gyrus			71	50	-46	2	6.27
R. Inferior Temporal Gyrus			68	56	-56	-16	6.45

L. Superior Temporal Gyrus / TPJ			43	-54	-44	18	6.54
(2) RHY: AO ∩ EXE (non-practised + practised)							
L. Pallidum			3632	-20	4	2	12.71
L. Insula Lobe*				-30	18	2	10.94
L. Inferior Frontal Gyrus (Pars Opercularis)*				-48	8	4	10.52
L. Precentral Gyrus*	Area 6	9.3		-42	-10	54	9.93
R. Cerebellum	Lobule VI	24.2	2317	32	-58	-26	16.09
L. Cerebellum*	Lobule VI	21.1		-32	-60	-24	14.58
L. Supplementary Motor Area (SMA)	Area 6	35.9	2221	-2	-2	60	17.01
R. Putamen			1002	20	10	0	8.88
L. Superior Temporal Gyrus / TPJ	IPC (PF)	9.8	924	-56	-42	20	12.98
L. Inferior Parietal Lobule*	hIP2	15.6		-48	-38	42	7.78
R. Precentral Gyrus	Area 6	19.2	769	50	0	42	9.00
R. Inferior Frontal Gyrus (Pars Opercularis)*	Area 44	32.2		52	12	20	7.99

R. Superior Temporal Gyrus / TPJ	IPC (PF)	27.9	612	62	-34	18	10.89
R. Cerebellum	Lobule VIIIa	26.4	375	28	-62	-50	13.97
R. Inferior Parietal Lobule	hIP1	35.3	266	36	-46	40	7.59

* Note that for some activation clusters we report more than the first maximum. Sub-maxima are indicated by an asterisk.

1001

Table 2. Conjunctions between sequence and rhythm tasks. Macroanatomical structure, cytoarchitectonical area (Areacyto), percent overlap of cluster with cytoarchitectonical area, cluster size in voxel, MNI coordinates (x, y, z), and maximum T value (Tmax) of the local maxima of the conjunctions between spatial sequences (SEQ) and rhythms (RHY), separately for action observation (AO) and execution (EXE) events, based on the activation differences between non-practised and practised patterns. Analyses included both groups. The significance level was set to $p < .001$, uncorrected. A cluster size of ≥ 70 contiguous voxels (560 mm^3) extended the threshold. MNI coordinates shown in bold indicate that the activation was also present at the higher threshold of $p < .05$, FWE-corrected, with a cluster size of ≥ 20 contiguous voxels (160 mm^3). Abbreviations: L. = left, R. = right.

				MNI coordinates			
Local maximum in macroanatomical structure	**$Area_{cyto}$**	**Percent overlap of cluster with cytoarchitectonical area**	**Cluster size (voxel)**	**x**	**y**	**z**	**T_{max}**
(1) AO: SEQ (non-practised > practised) ∩ RHY (non-practised > practised)							
R. Inferior Frontal Gyrus (Pars Opercularis)	Area 44	49.8	832	**50**	**10**	**14**	**7.19**
R. Precentral Gyrus*				40	2	34	4.61
R. Middle Temporal Gyrus			712	48	-44	8	4.48
L. Supplementary Motor Area (SMA)			596	**-6**	**12**	**48**	**5.59**
L. Inferior Frontal Gyrus (Pars Opercularis)	Area 44	26.4	442	-46	12	20	4.43
L. Precentral Gyrus*				-44	-2	36	4.32
L. Middle Temporal Gyrus			264	**-50**	**-50**	**8**	**5.53**
R. Inferior Parietal Lobule	IPC (PFt)	42.9	248	48	-34	46	4.82

L. Middle Temporal Gyrus			229	-46	-66	6	4.23
R. Inferior Frontal Gyrus (Pars Triangularis)	Area 45	31.5	93	50	36	10	4.20
R. Insula Lobe			82	30	24	-4	3.51
(2) EXE: SEQ (non-practised > practised) ∩ RHY (non-practised > practised)							
R. Anterior Cingulate Cortex			2745	**4**	**28**	**26**	**7.48**
L. Anterior Cingulate Cortex*				**-2**	**26**	**28**	**7.36**
L. Middle Cingulate Cortex*				**-4**	**26**	**32**	**7.31**
L. Supplementary Motor Area (SMA)*				**0**	**12**	**54**	**7.01**
L. Insula Lobe			1677	**-28**	**22**	**-4**	**7.53**
L. Inferior Frontal Gyrus (Pars Triangularis)*	Area 45	9.4		**-52**	**18**	**20**	**5.65**
L. Inferior Frontal Gyrus (Pars Opercularis)*	Area 44	16.7		**-46**	**12**	**6**	**5.20**
R. Insula Lobe			1132	**34**	**22**	**-2**	**5.47**
R. Inferior Frontal Gyrus (Pars Triangularis)*				46	28	28	4.89
R. Middle Frontal Gyrus*				44	40	20	4.56

L. Middle Frontal Gyrus	123	-30	40	14	3.99

* Note that for some activation clusters we report more than the first maximum. Sub-maxima are indicated by an asterisk.

Figure captions

Colour reproduction of Figures 3, 4, 5, and 6 is necessary on the web as well as in print.

Figure 1. Experimental design. Participants were tested on practised as well as non-practised patterns of spatial sequences (SEQ) and rhythms (RHY) in three presentation conditions: Action Observation (AO: video observation followed by rest), Motor Imagery (MI: video observation followed by motor imagery), and Action Execution (EXE: video observation followed by imitative execution). All conditions of the 3 x 2 x 2 experimental design (AO / MI / EXE, SEQ / RHY, practised / non-practised) were presented in pseudo-randomized order. Each trial started with a fixation cue (white square) in the center of the screen for a duration of 1 s to direct particpants' attention. The cue was followed by a 4.7 s long video clip showing either a spatial sequence or a rhythm. During video observation participants were unaware about the subsequent task instruction. In the AO condition, the screen turned black after the video presentation, which indicated a rest period that ranged between 3 and 14 s and served as baseline. In the MI condition, video observation was followed by a task cue (red square) lasting between 1 and 3.4 s. This indicated that a large grey square, of the same size as the video clips, would soon appear which then served as the go-signal for motor imagery of the previously observed pattern. After 4.7 s, a black screen appeared for a duration of 5.9 s, which served as rest baseline. In the EXE condition, a different task cue (green cross) indicated overt imitation. Due to the jittered task cue duration, the total duration of MI and EXE trials ranged between 17.3 s and 19.7 s.

Figure 2. Behavioural data. The imitation performance in the execution trials was analysed by means of a sliding window over three consecutive responses ('triplets'), where six correct triplets indicate correct imitation of the eight spatial positions or temporal intervals. For statistical results, see text.

Figure 3. Task networks for sequence and rhythm imitation. Conjunction analyses between action observation and execution separately for spatial sequences (SEQ: green) and rhythms (RHY: red). Analyses included both groups as well as non-practised and practised patterns. Images were thresholded at $p < .05$, FEW-corrected for the whole brain volume with an extent of $k = 20$ voxel (160 mm^3), superimposed on left, top, and right views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/).

Figure 4. Practice effects. Activation differences between non-practised and practised patterns, separately for action observation, motor imagery, and execution events, and for spatial sequences (SEQ: green) and rhythms (RHY: red). Analyses included both groups. Images were thresholded at $p < .001$, uncorrected with an extent of $k = 70$ voxel (560 mm^3), superimposed on left, top, and right views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/). *Activation decreases with practice*. *AO / SEQ:* bilateral occipital and posterior temporal regions, SPL, IPL, bilateral precentral gyrus, pars opercularis of IFG (Area 44), right pars triangularis of IFG (Area 45), SMA, middle cingulate cortex, and right insular cortex. *AO / RHY*: bilateral superior temporal gyrus / TPJ, pars opercularis and pars triangularis of IFG (Area 44 and 45,

resp.), SMA, as well as middle and inferior temporal regions, right IPL, left parietal operculum, precentral gyrus, left insula, and subcortically putamen and cerebellum bilaterally. *MI / SEQ*: bilateral IPL, SMA, bilateral IFG and postcentral gyrus, the left insula, left anterior and middle cingulate cortex, and middle frontal gyrus (MFG) bilaterally. *MI / RHY*: right IPL and cerebellum. *EXE / SEQ:* SMA, precentral gyrus extending to pars opercularis of the IFG, bilateral MFG, anterior and middle cingulate cortex, insula, bilateral IPL, and cerebellum. *EXE / RHY*: SMA, bilateral pars opercularis and pars triangularis of IFG, right MFG, anterior and middle cingulate cortex, bilateral insula, and two small activation clusters in the right cerebellum and left pallidum and thalamus. *Activation increases with practice. AO / SEQ*: merely midline structures showed activation increases, namely bilateral cingulate cortex and precuneus, as well as left angular gyrus, left hippocampus, left cerebellum, and bilateral basal ganglia. *AO / RHY*: left occipital cortex, angular gyrus, and precuneus. *MI:* no activation increases with practice for either task. *EXE / SEQ*: middle and posterior cingulate cortex, left SPL, right parietal operculum (OP1), and subcortically amygdala, putamen, and right cerebellum. *EXE / RHY*: right middle cingulate cortex, right parietal operculum (OP1), bilateral IPL, and right amygdala and putamen.

Figure 5. Conjunctions between sequence and rhythm tasks. Conjunction between spatial sequence and rhythm imitation tasks, separately for action observation and execution events, based on the activation differences between non-practised and practised patterns across musicians and non-musicians. Images with red colour range were thresholded at $p < .001$, uncorrected with an extent of $k = 70$ voxel (560 mm^3), and images with yellow colour range were thresholded at $p < .05$, FWE-corrected with an extent of $k = 20$ voxel (160 mm^3). All images were superimposed on left, top,

right, and midsaggital views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/).

Figure 6. Practice effects in non-musicians and musicians. Differences between non-practised and practised patterns in each participant group, separately for sequences (SEQ: green) and rhythms (RHY: red), and for action observation and execution events. Images were thresholded at $p < .001$, uncorrected with an extent of $k = 70$ voxel (560 mm^3), superimposed on left, top, and right views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/).

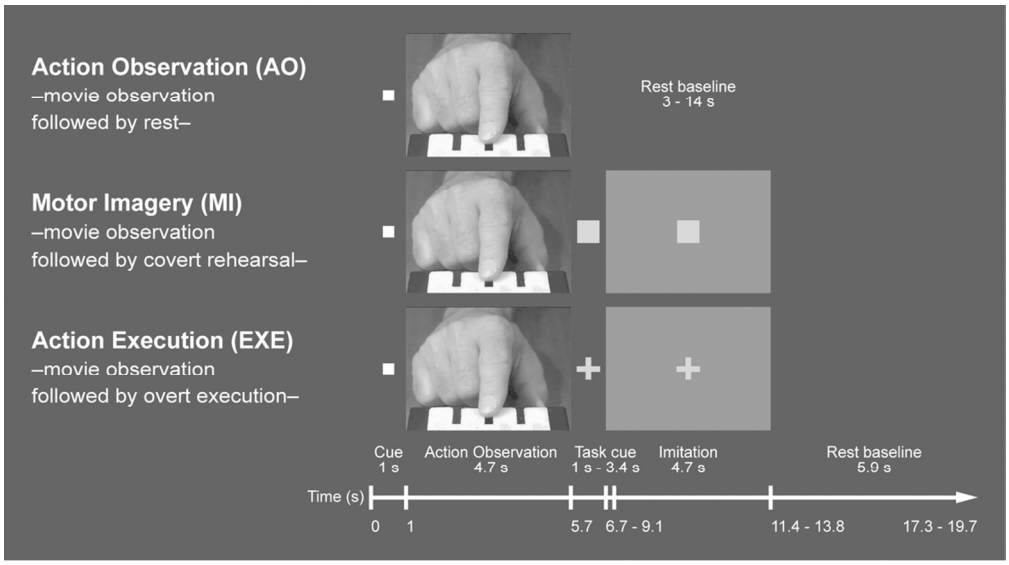

Figure 1. Experimental design. Participants were tested on practised as well as non-practised patterns of spatial sequences (SEQ) and rhythms (RHY) in three presentation conditions: Action Observation (AO: video observation followed by rest), Motor Imagery (MI: video observation followed by motor imagery), and Action Execution (EXE: video observation followed by imitative execution). All conditions of the 3 x 2 x 2 experimental design (AO / MI / EXE, SEQ / RHY, practised / non-practised) were presented in pseudo-randomized order. Each trial started with a fixation cue (white square) in the center of the screen for a duration of 1 s to direct particpants' attention. The cue was followed by a 4.7 s long video clip showing either a spatial sequence or a rhythm. During video observation participants were unaware about the subsequent task instruction. In the AO condition, the screen turned black after the video presentation, which indicated a rest period that ranged between 3 and 14 s and served as baseline. In the MI condition, video observation was followed by a task cue (red square) lasting between 1 and 3.4 s. This indicated that a large grey square, of the same size as the video clips, would soon appear which then served as the go-signal for motor imagery of the previously observed pattern. After 4.7 s, a black screen appeared for a duration of 5.9 s, which served as rest baseline. In the EXE condition, a different task cue (green cross) indicated overt imitation. Due to the jittered task cue duration, the total duration of MI and EXE trials ranged between 17.3 s and 19.7 s.

< please enter Figure 1 about
99x55mm (300 x 300 DPI)

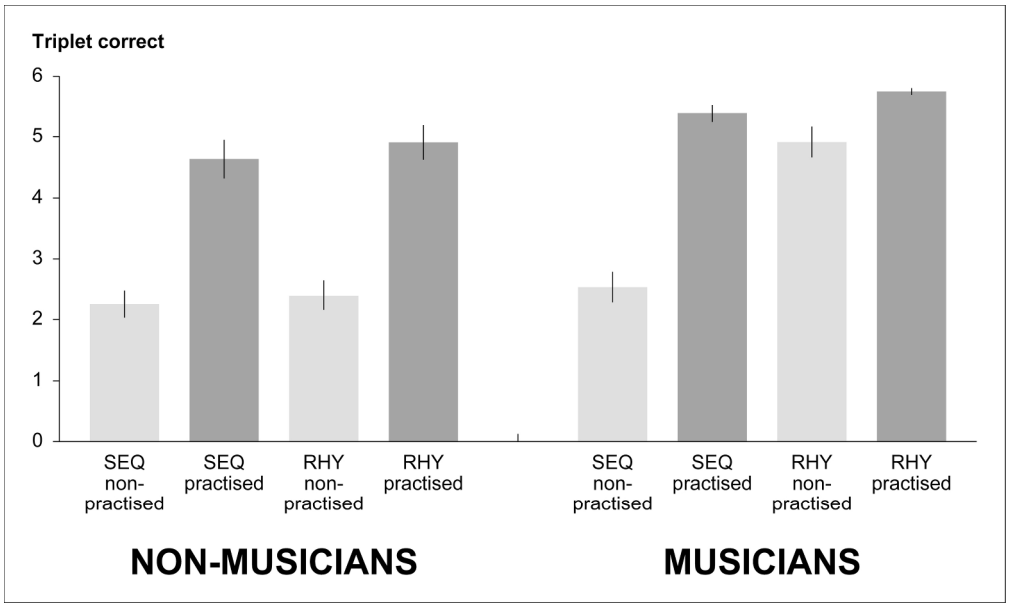

Figure 2. Behavioural data. The imitation performance in the execution trials was analysed by means of a sliding window over three consecutive responses ('triplets'), where six correct triplets indicate correct imitation of the eight spatial positions or temporal intervals. For statistical results, see text.
< please enter Figure 2 about
107x64mm (600 x 600 DPI)

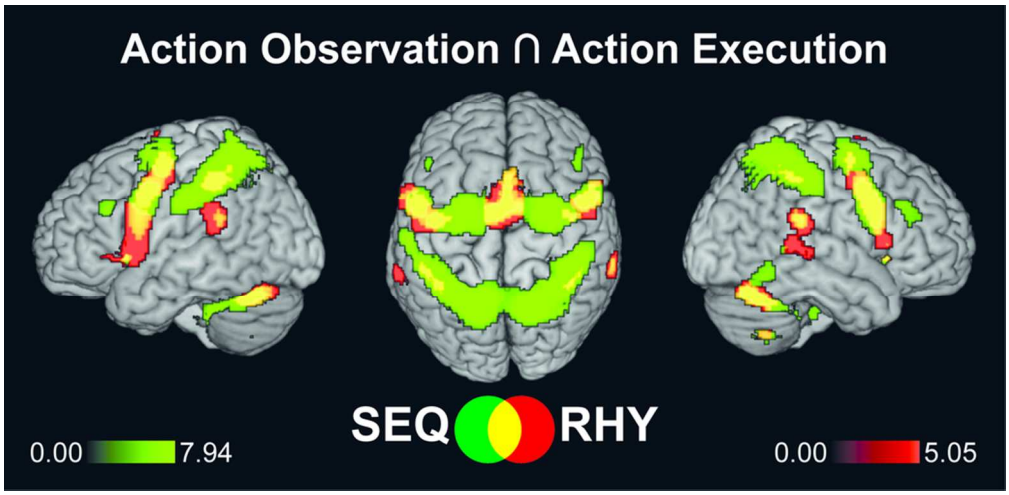

Figure 3. Task networks for sequence and rhythm imitation. Conjunction analyses between action observation and execution separately for spatial sequences (SEQ: green) and rhythms (RHY: red). Analyses included both groups as well as non-practised and practised patterns. Images were thresholded at $p < .05$, FEW-corrected for the whole brain volume with an extent of $k = 20$ voxel (160 mm3), superimposed on left, top, and right views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/).

< please enter Figure 3 about
87x42mm (300 x 300 DPI)

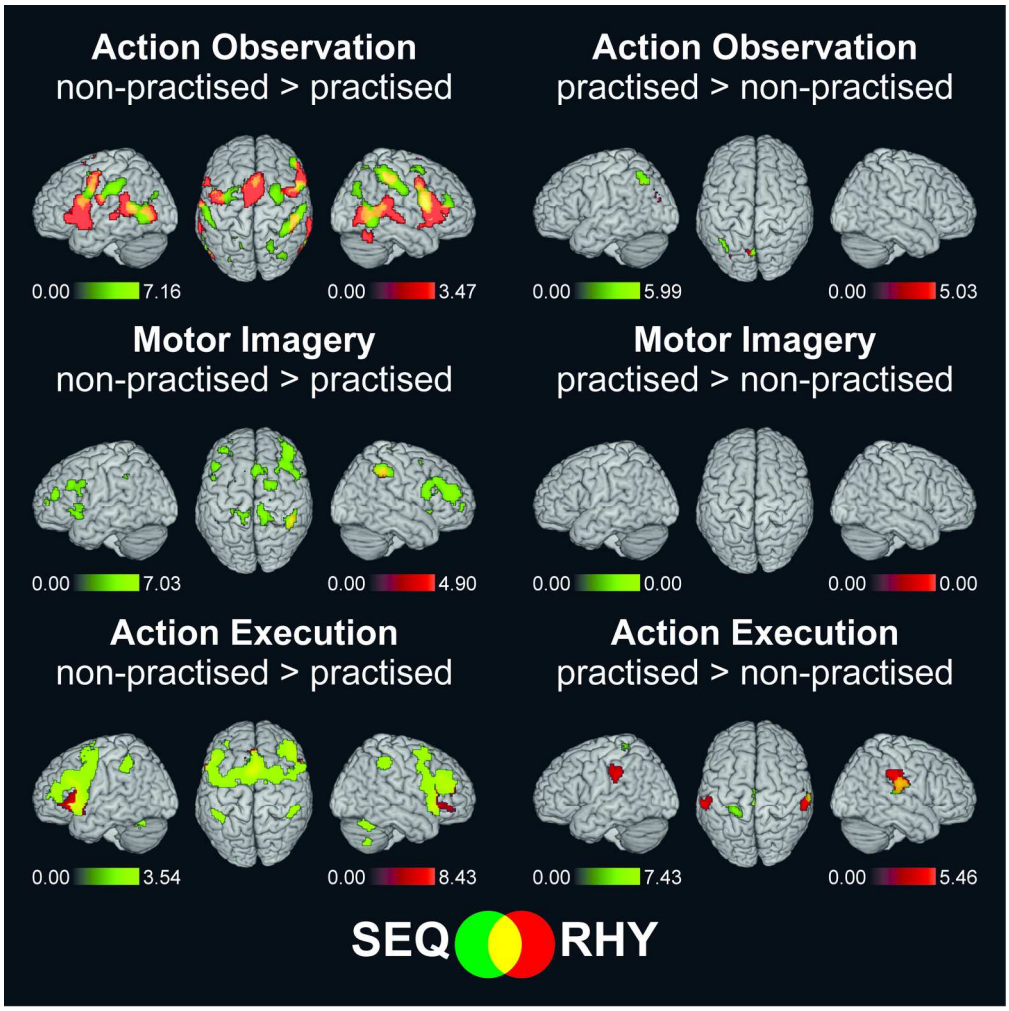

Figure 4. Practice effects. Activation differences between non-practised and practised patterns, separately for action observation, motor imagery, and execution events, and for spatial sequences (SEQ: green) and rhythms (RHY: red). Analyses included both groups. Images were thresholded at $p < .001$, uncorrected with an extent of $k = 70$ voxel (560 mm3), superimposed on left, top, and right views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/). Activation decreases with practice. AO / SEQ: bilateral occipital and posterior temporal regions, SPL, IPL, bilateral precentral gyrus, pars opercularis of IFG (Area 44), right pars triangularis of IFG (Area 45), SMA, middle cingulate cortex, and right insular cortex. AO / RHY: bilateral superior temporal gyrus / TPJ, pars opercularis and pars triangularis of IFG (Area 44 and 45, resp.), SMA, as well as middle and inferior temporal regions, right IPL, left parietal operculum, precentral gyrus, left insula, and subcortically putamen and cerebellum bilaterally. MI / SEQ: bilateral IPL, SMA, bilateral IFG and postcentral gyrus, the left insula, left anterior and middle cingulate cortex, and middle frontal gyrus (MFG) bilaterally. MI / RHY: right IPL and cerebellum. EXE / SEQ: SMA, precentral gyrus extending to pars opercularis of the IFG, bilateral MFG, anterior and middle cingulate cortex, insula, bilateral IPL, and cerebellum. EXE / RHY: SMA, bilateral pars opercularis and pars triangularis of IFG, right MFG, anterior and middle cingulate cortex, bilateral insula, and two small activation clusters in the right cerebellum and left pallidum and thalamus. Activation increases with practice. AO / SEQ: merely midline structures showed activation increases, namely bilateral cingulate cortex and precuneus, as well as left angular gyrus, left hippocampus, left cerebellum, and bilateral basal ganglia. AO / RHY: left occipital cortex, angular gyrus, and precuneus. MI: no activation increases with practice for either task. EXE / SEQ: middle and posterior cingulate cortex, left SPL, right parietal operculum (OP1), and subcortically amygdala, putamen, and right cerebellum. EXE / RHY: right middle cingulate

cortex, right parietal operculum (OP1), bilateral IPL, and right amygdala and putamen.
< please enter Figure 4 about
180x180mm (300 x 300 DPI)

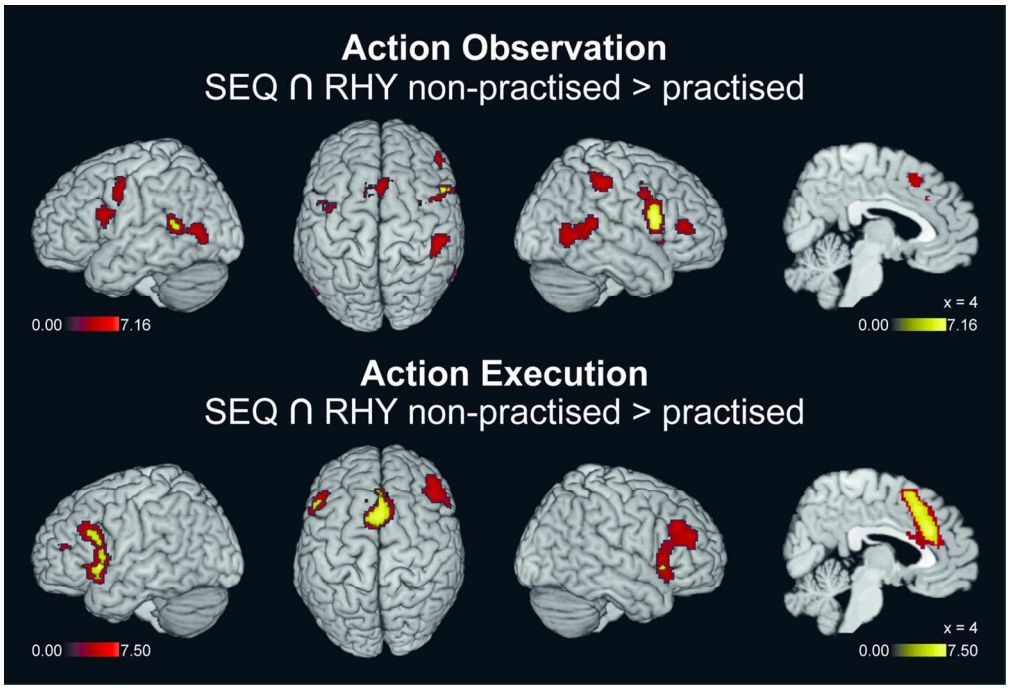

Figure 5. Conjunctions between sequence and rhythm tasks. Conjunction between spatial sequence and rhythm imitation tasks, separately for action observation and execution events, based on the activation differences between non-practised and practised patterns across musicians and non-musicians. Images with red colour range were thresholded at $p < .001$, uncorrected with an extent of $k = 70$ voxel (560 mm3), and images with yellow colour range were thresholded at $p < .05$, FWE-corrected with an extent of $k = 20$ voxel (160 mm3). All images were superimposed on left, top, right, and midsaggital views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/).
< please enter Figure 5 about
122x82mm (300 x 300 DPI)

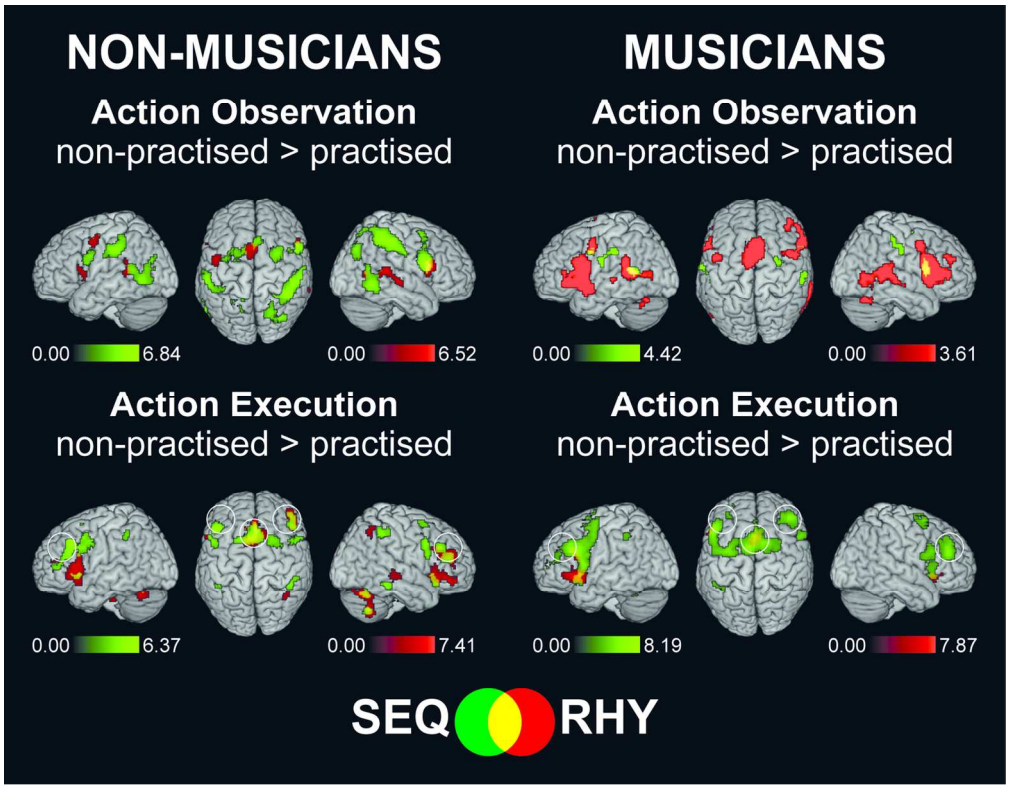

Figure 6. Practice effects in non-musicians and musicians. Differences between non-practised and practised patterns in each participant group, separately for sequences (SEQ: green) and rhythms (RHY: red), and for action observation and execution events. Images were thresholded at $p < .001$, uncorrected with an extent of $k = 70$ voxel (560 mm3), superimposed on left, top, and right views of the volume rendered MNI template using the software MRIcron Version 6/2013 (http://www.nitrc.org/projects/mricron/).

< please enter Figure 6 about
139x108mm (300 x 300 DPI)