CrossMark

ORIGINAL ARTICLE

Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions

E. E. Hecht · D. A. Gutman · N. Khreisheh ·
S. V. Taylor · J. Kilner · A. A. Faisal ·
B. A. Bradley · T. Chaminade · D. Stout

Abstract Human ancestors first modified stones into tools 2.6 million years ago, initiating a cascading increase in technological complexity that continues today. A parallel trend of brain expansion during the Paleolithic has motivated over 100 years of theorizing linking stone toolmaking and human brain evolution, but empirical support remains limited. Our study provides the first direct experimental evidence identifying likely neuroanatomical targets of natural selection acting on toolmaking ability. Subjects received MRI and DTI scans before, during, and after a 2-year Paleolithic toolmaking training program. White matter fractional anisotropy (FA) showed changes in branches of the superior longitudinal fasciculus leading into left supramarginal gyrus, bilateral ventral precentral gyri, and right inferior frontal gyrus pars triangularis. FA increased from Scan 1–2, a period of intense training, and decreased from Scan 2–3, a period of reduced training. Voxel-based morphometry found a similar trend toward

E. E. Hecht · D. Stout
Department of Anthropology, Emory University, 1557 Dickey Drive, Room 114, Atlanta, GA 30322, USA
e-mail: dwstout@emory.edu

E. E. Hecht (✉)
Department of Psychology, Center for Behavioral Neuroscience, Georgia State University, Atlanta, GA, USA
e-mail: ehecht@gsu.edu

D. A. Gutman
Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, USA
e-mail: dagutman@emory.edu

D. A. Gutman
Department of Psychiatry and Behavioral Science, School of Medicine, Emory University, Atlanta, USA

D. A. Gutman
Center for Comprehensive Informatics, Emory University, Psychology Building, Suite 563, 36 Eagle Row, Atlanta, GA 30322, USA

N. Khreisheh · B. A. Bradley
Department of Archaeology, University of Exeter, Laver Building, North Park Road, Exeter EX4 4QE, UK
e-mail: N.N.Khreisheh@ex.ac.uk

B. A. Bradley
e-mail: B.A.Bradley@exeter.ac.uk

S. V. Taylor · A. A. Faisal
Department of Bioengineering, Imperial College London, Royal School of Mines Building, Prince Consort Rd, SW7 2AZ London, UK
e-mail: s.taylor10@imperial.ac.uk

A. A. Faisal
e-mail: a.faisal@imperial.ac.uk

J. Kilner
The Sobell Department of Movement Neuroscience and Motor Disorders, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK
e-mail: j.kilner@ucl.ac.uk

A. A. Faisal
Department of Bioengineering and Computing, Imperial College London, Royal School of Mines Building, Prince Consort Rd, SW7 2AZ London, UK

A. A. Faisal
MRC Clinical Sciences Center, London, UK

T. Chaminade
Institut de Neurosciences de la Timone UMR 7289, Aix Marseille Université, CNRS, 13385 Marseille, France
e-mail: tchamina@gmail.com

gray matter expansion in the left supramarginal gyrus from Scan 1–2 and a reversal of this effect from Scan 2–3. FA changes correlated with training hours and with motor performance, and probabilistic tractography confirmed that white matter changes projected to gray matter changes and to regions that activate during Paleolithic toolmaking. These results show that acquisition of Paleolithic toolmaking skills elicits structural remodeling of recently evolved brain regions supporting human tool use, providing a mechanistic link between stone toolmaking and human brain evolution. These regions participate not only in toolmaking, but also in other complex functions including action planning and language, in keeping with the hypothesized co-evolution of these functions.

Keywords Diffusion tensor imaging · Plasticity · Tool use · Brain evolution · Language · Superior longitudinal fasciculus

> …*first labour, after it and then with it speech—these are the two most essential stimuli under the influence of which the brain of the ape gradually changed into that of man.*
> Friedrich Engels, *The Part Played by Labour in the Transition from Ape to Man*, 1876.

Introduction

Many species make and use tools, but humans are distinguished by the extent and complexity of their technological behavior. It is likely that the neural prerequisites for this uniquely human facility evolved during the Paleolithic (2,600,000–10,000 years ago), a period which witnessed the origin (Semaw et al. 2003) and evolution (Stout 2011) of stone toolmaking. During the same time period, the hominin brain underwent a roughly threefold increase from ape to human proportions (Holloway et al. 2004a, b). Over 100 years of theorizing has linked Paleolithic toolmaking to human brain evolution (Oakley 1949; Holloway 1967; Engels 2003 [1876]), but empirical evidence bearing on the exact nature of this interaction has remained scant. In order to identify likely neuroanatomical targets of selection acting on toolmaking ability, we replicated the skill learning challenges faced by our early toolmaking ancestors and measured structural effects on the brain. This experiment builds on previous comparative work to test the hypothesis that stone toolmaking can elicit plastic structural responses in evolutionarily relevant brain structures. Such phenotypic plasticity enables learning to facilitate evolutionary adaptation via what is commonly termed the "Baldwin effect" (Weber and Depew 2003; Bateson 2004), potentially

providing a mechanistic link between Paleolithic toolmaking and human brain evolution.

Our experiment builds on two lines of research into human brain evolution: comparative studies with nonhuman primates and experimental "neuroarchaeological" studies with modern humans replicating Paleolithic technology. These methods have complementary strengths and weaknesses, and are best used in conjunction. Comparative studies provide evidence of differences in brain structure and function between extant species that can be used to identify evolutionary changes specific to the human lineage. Comparative studies cannot, however, reveal the more specific timing and behavioral context of adaptations that occurred since the last common ancestor of species under consideration [e.g., chimpanzees and human LCA, ca. 7–8 million years ago (Langergraber et al. 2012)]. The archeological record provides an underutilized resource to fill this gap. Archeological evidence allows the detailed reconstruction of ancient human behaviors, which may then be studied using experimental neuroscience methods. A limitation of this approach is that it must use evidence from modern human subjects to support inferences about brain responses in pre-modern hominins. It is thus important that neuroarchaeological hypotheses and interpretations be constrained by comparative evidence. Here we focus on shared brain systems known to support tool use in humans and nonhuman primates, and which also show additional human specializations for tool use. The substantial homology of these systems across nonhuman primates and modern humans makes it unlikely that ancestral hominins displayed a uniquely divergent overall functional organization. Moreover, experimental results linking particular, archeologically attested toolmaking behaviors to modern human specializations within this network support inferences regarding the likely timing and behavioral context for the emergence of these specializations.

Comparative studies with nonhuman primates point toward human specializations in frontoparietal tool-use networks. For example, humans activate a region of anterior supramarginal gyrus during the observation of tool use actions, while macaque monkeys fail to activate the homologous region (Peeters et al. 2009). This region responds to the rigid kinematics of tools as opposed to the biological motion of hands (Peeters et al. 2013) and, together with enhanced parietal contributions to 3-dimensional form-from-motion perception (Vanduffel et al. 2002), has been proposed to support uniquely human tool use capacities (Peeters et al. 2009; Orban and Rizzolatti 2012). While macaques do not naturally use tools, experimentally tool-trained macaques show increases in gray matter in superior temporal sulcus, secondary somatosensory cortex, and intraparietal sulcus (Quallo et al. 2009). Compared to humans, macaques also have relatively more

prefrontal activation during object perception (Denys et al. 2004), and relatively more frontal and less parietal activation during the perception of grasping actions (Nelissen et al. 2005). A comparison between human and chimpanzee brain activation during the observation of object-directed grasping found relatively greater human activation in ventral premotor cortex, inferior parietal cortex, and inferotemporal cortex, with human activation being generally more posterior than chimpanzee activation (Hecht et al. 2013b). Anatomically, lateral frontal, parietal and temporal association cortices are among the most volumetrically expanded portions of the human brain (Hill et al. 2010; Avants et al. 2006). This parallels comparative studies of frontoparietal white matter connectivity, which have reported relatively greater human connectivity with parietal and temporal regions (Rilling et al. 2008; Hecht et al. 2013a) in tracts that may be involved in human tool use (Ramayya et al. 2010). We have argued (Hecht et al. 2013b) that this indicates a greater contribution for "bottom-up" representation of observed action in the human brain, with more neural resources devoted to processing the details of objects and movements. Because action understanding is fundamental to tool use, social learning, and cumulative culture, we suggest that the evolution of these functions may have involved modifications of frontoparietal interactions.

Neuroarchaeological methods have previously been applied in functional studies of Paleolithic toolmaking using FDG-PET (Stout and Chaminade 2007; Stout et al. 2008), fMRI (Stout et al. 2011), and transcranial Doppler ultrasound (Uomini and Meyer 2013). Echoing the comparative evidence of frontoparietal elaboration in human evolution, experimental results have identified a distributed, bilateral frontoparietal network supporting stone toolmaking in modern humans. Some of these activation loci are plotted in Fig. 8. Early toolmaking ("Oldowan", ca. 2.6 million years ago) activates ventral premotor and parietal regions identified by comparative studies as likely foci of human perceptual-motor specializations for tool use (see above). More advanced toolmaking ("Acheulean handaxe", ca. 500,000 years ago) generates increased activation of inferior prefrontal cortex (Stout et al. 2008, 2011), another major locus of structural and functional change in human brain evolution (Falk 1983; Holloway et al. 2004a, b; Schenker et al. 2008; Teffer and Semendeferi 2012). This supports an evolutionary scenario in which perceptual-motor adaptations enabled the initial stages of human technological evolution, whereas later developments were dependent on enhanced cognitive control (Faisal et al. 2010; Stout 2010). Furthermore, observed overlap in functional neuroanatomy (Stout and Chaminade 2012) and the time-course of hemodynamic responses (Uomini and Meyer 2013) between stone toolmaking and language has renewed

support to longstanding hypotheses of an evolutionary connection between these two distinctive human capacities (Holloway 1967). In keeping with broader "gestural origin" hypotheses of language evolution (e.g., Arbib 2005, Corballis 2002), it has been proposed that adaptations for toolmaking were evolutionarily co-opted ["exapted" (Gould and Vrba 1982)] at various points in human evolution to support proto-linguistic communicative behaviors (Pulvermüller and Fadiga 2010; Greenfield 1991; Stout and Chaminade 2012).

Together, the comparative (Peeters et al. 2009, 2013; Vanduffel et al. 2002; Orban and Rizzolatti 2012; Denys et al. 2004; Nelissen et al. 2005; Hecht et al. 2013a, b; Rilling et al. 2008; Ramayya et al. 2010) and neuroarchaeological (Stout and Chaminade 2007; Stout et al. 2008, 2011; Uomini and Meyer 2013; Faisal et al. 2010) studies reviewed here suggest that selection pressure for toolmaking ability may have influenced the evolution of human inferior frontoparietal regions. However, task-related functional activations cannot show that stone toolmaking actually generates pressure for structural changes of the kind (e.g., frontoparietal expansion and increased connectivity) indicated by the comparative evidence. To test this, we replicated the learning challenges faced by our early toolmaking ancestors and measured structural effects on the brain. Six subjects underwent an intensive, 2-year training program in a variety of archeologically attested Paleolithic toolmaking methods (Fig. 1). Structural MRI and DTI scans were collected at the start (Time 1), during (Time 2), and at the end (Time 3) of training. Individual training time and sensorimotor performance data for these time points were also collected, in order to directly test the correlation of observed structural changes with stone toolmaking training (cf. Thomas and Baker 2013). Stone toolmaking is a complex, real-world craft and a simple, objective method for quantifying variation in overall toolmaking "skill" does not exist. We chose to measure one key aspect of toolmaking skill, the accuracy of percussion strikes, using a controlled experimental task. Previous research has shown that control over percussive accuracy is critical to the planning and control of toolmaking outcomes and varies with toolmaking experience (Nonaka et al. 2010).

We expected that, if stone toolmaking does indeed place acute demands on the anatomical structure of frontoparietal cortex, this should be reflected in measurable changes in gray and white matter in this region and that the magnitude of these changes should correlate with training time and behavioral measures. These are conservative predictions because it is likely that the evolved frontoparietal cortices of our modern human subjects are already structurally well-adapted to the skilled use of tools, and thus less likely to show additional plastic response to stone toolmaking practice. If toolmaking nevertheless generates a measurable
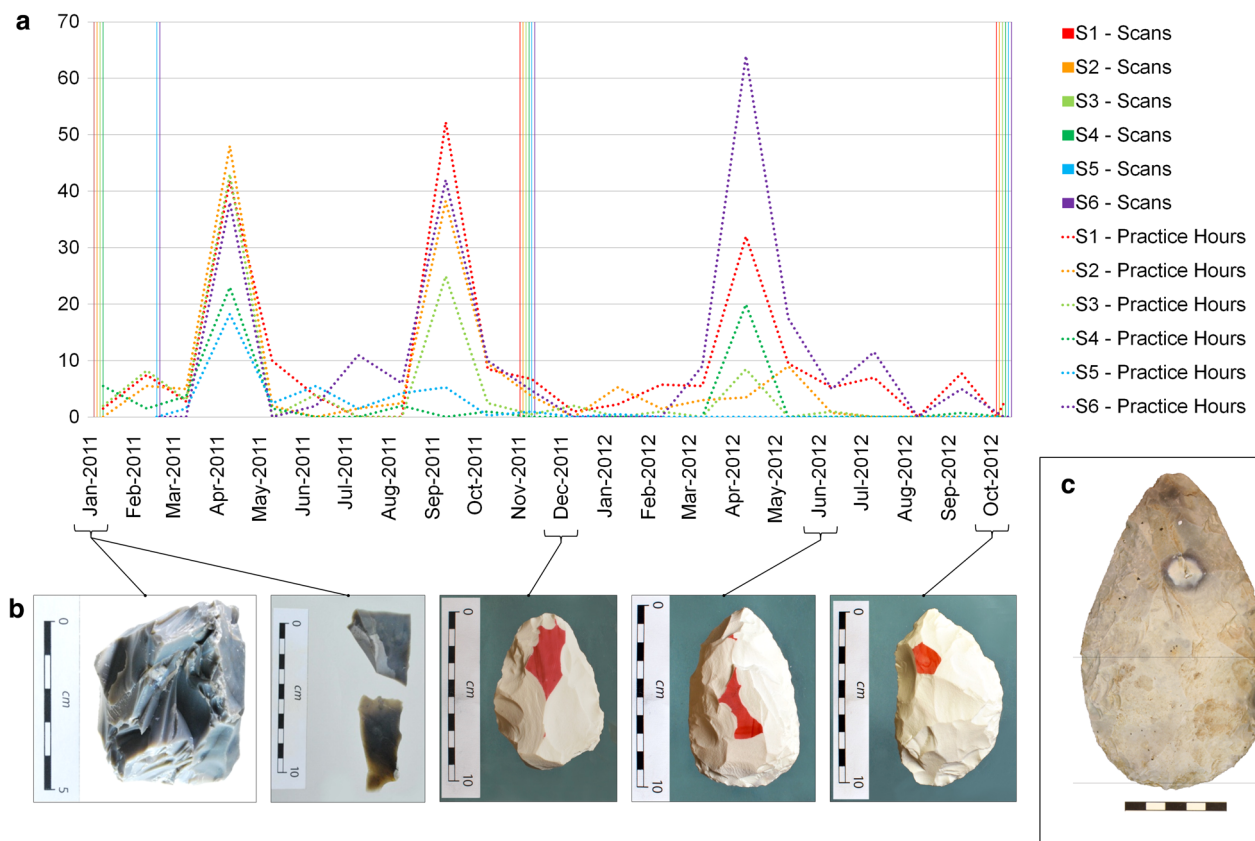
Fig. 1 Intensive 2-year training program in Paleolithic stone tool-making. **a** Training included direct instruction and individual practice, both on site and during intensive field trips (04/2011, 09/2011, 04/2012). **b** Stone tools produced by a representative subject at T1, T2, and T3. The last three tools were created using porcelain which was coated with *red paint* in order to facilitate recognition of the original surface (Khreisheh et al. 2013). **c** An actual Paleolithic stone tool (Acheulean handaxe, ∼500,000 years old)

response in modern human subjects, this provides a strong indication that comparable toolmaking by ancestral hominins would have stressed homologous substrates to an equal or greater extent.

Materials and methods

Subjects

Subjects were recruited from the undergraduate and post-graduate programs in Archaeology at Exeter University. Subjects were ages 18–25 at the time the first scan was collected, five males and one female. All were right-handed by self-report and had no neurological or psychiatric illness. All subjects consented to participate and the study was approved by the Ethics Committee at Exeter University.

Toolmaking training

Paleolithic stone toolmaking involves striking a stone "core" with a "percussor" of bone, antler, or stone to detach carefully controlled chips and incrementally achieve various design goals. This demanding task requires precise sensorimotor coordination and strategic action sequence planning. Training occurred on site at Exeter University and during intensive field trips to the USA, France, and Denmark. Subjects were trained by Bruce Bradley, Ph.D., Director of the Experimental Archaeology Masters Programme at Exeter University, and an expert stone tool-maker with decades of experience teaching Paleolithic technologies. Training included instruction, coaching, and demonstration as well as independent practice, which was recorded by subjects in a log book. The Paleolithic tool-making methods learned by subjects included: (1) basic flake production, comparable to the earliest known (Oldowan) tools of *Homo habilis* 2.6–1.5 million years ago (mya); (2) "Handaxe" making, comparable to the Acheulean tools of *Homo erectus* and *Homo heidelbergensis* 1.7–0.25 mya; and (3) "prepared core" flake production, comparable to the Levallois tools of Neanderthals and early *Homo sapiens* <0.25 mya. Training was naturalistic and self-paced, resulting in substantial variation across subjects in the duration, intensity, and content of practice.

Paleolithic toolmaking occurred over a vast time period and across many millions of square miles, and naturally

encompasses a great deal of variation that could not be included in the methods learned by our subjects. The methods we did select are considered broadly representative of Early and Middle Paleolithic technology, whereas details of the production techniques employed closely match those documented in specific archeological collections (e.g., Stout et al. 2014, Bradley and Sampson 1986). This gives us confidence that our training protocol was both generally representative and specifically accurate in re-creating learning challenges actually faced by Paleolithic toolmakers.

Motor performance

Motor testing took place at the Brain and Behaviour Lab, Department of Bioengineering and Department of Computing, Imperial College London, on the same days as MR image acquisition. Due to scheduling constraints performance data were collected for only five (four male) of the six subjects. All subjects provided written informed consent, and experiments were carried out in accordance with institutional guidelines. A local ethics committee approved the experimental protocols.

Subjects held in their right hand an object (250 g, hand-sized cylinder) with an embedded Liberty motion tracking node sampled at 240 Hz, calibrated to sub-millimeter precision and tracked by a Polhemus Liberty electromagnetic tracking system [POLHEMUS, Colchester (VT)]. The measurement markers of this system recorded horizontal, vertical and depth position of the marker in a calibrated reference coordinate system (see Faisal et al. 2010). The coordinate system is aligned veridical with an augmented reality projection system linked to a computer controlling the experiment and processing the position information in real time (akin to Faisal and Wolpert 2009).

Subjects were instructed to strike from three different strike positions towards a fixed target point. The start points were located on a line 7.5, 24, and 40 cm away from the target point. The spatial position of the start point and target were located on an imaginary line, and mimicked the motion of percussion strikes during natural knapping. Start positions were randomly selected before each trial. Subjects had to move the marker to the start position before being able to initiate the strike. Subjects could initiate a strike at any time of their choosing, but once movement was initiated subjects had a time constraint of 200 ms to hit the target, or had to repeat the trial (classified as miss). Invalid trails (e.g., movement not executed in time window) were discarded. Subjects first performed 500 trials with a circular target of 2 cm, then 500 trials with a circular target of 1 cm, for a total average of 920 ± 31.7 (range 845–991) valid trials per session. We calculated accuracy (proportion of valid trials passing through target) per

session and tested for correlation between changes in accuracy and changes in fractional anisotropy (FA, see Sect. "Tract-based spatial statistics") using Pearson's r correlation coefficient.

Image acquisition and preprocessing

Imaging took place at the Wellcome Department of Imaging Neuroscience in London. All subjects consented and the research was approved by the National Hospital for Neurology and Neurosurgery and Institute of Neurology Joint Research Ethics Committee. 61-direction DTI images were acquired on a Siemens Trio 3.0 T scanner with a voxel size of 1.7 mm^3. Seven volumes were acquired with no diffusion weighting (B0s); these images provide a "baseline" from which diffusion magnitude and directionality can be measured. The FSL software package (Smith et al. 2004; Woolrich et al. 2009; Jenkinson et al. 2012) was used for image processing and analysis. The first B0 image from each scan was discarded to allow for equipment warm-up and the remaining 6 were averaged. FSL's brain extraction tool (BET) (Smith 2002) was applied to the averaged B0s, which were then aligned to template space using a nonlinear registration algorithm (FNIRT) (Andersson et al. 2007) with an initial affine registration (FLIRT) (Jenkinson and Smith 2001; Jenkinson et al. 2002). FSL's eddy correction algorithm (EDDY) was used to correct distortion caused by eddy currents. DTIFIT, part of FSL's FDT package (Behrens et al. 2003) was used to fit the raw diffusion data to a tensor model at each voxel to obtain diffusion scalars, including fractional anisotropy. Dates for each subject's scans at Time 1, 2, and 3 are listed in Table 1.

Registration to template space

Since the goal of the present study was to compare scans within subjects across time, we used an approach similar to a previous longitudinal DTI study (Engvig et al. 2012) in order to ensure accurate within-subject alignment across time points despite potential differences in the subject's

Table 1 Time points for each subject's MRI and DTI scans, given in days from the beginning of the study

Subject	Days from Scan 1 to Scan 2	Days from Scan 1 to Scan 3
Subject 1	308	644
Subject 2	307	644
Subject 3	306	644
Subject 4	306	644
Subject 5	266	603
Subject 6	265	604

position in the scanner or in equipment calibration. First, for each time point, each subject's fractional anisotropy (FA) image, which is a scalar measure of diffusion magnitude at each voxel, was linearly registered to the FMRIB 1 mm^3 FA template, which is a standard T1-weighted MRI template formed by averaging the scans of 152 subjects. Next, FSL's *midtrans* algorithm was used to find the transformation matrix from a geometrically intermediate space of each subject's 3 FA images to the FA template (their "midspace"). This matrix was inverted and applied to the FA template to transform the template into each subject's FA midspace. Each subject's 3 native-space FA images were then linearly registered to their own midspace template. These images were fed into the FSL software package's standard TBSS pipeline (see below), which includes nonlinear registration to a common template space. For probabilistic tractography, each subject's midspace FA template was nonlinearly aligned to the MNI-space FMRIB 1 mm^3 FA template. This warp was inverted and combined with the linear midspace-to-native-space transformation matrix in order to produce the MNI space-to-native-space transformation matrix for each subject, which was used to apply standard-space regions of interest to native diffusion space (Engvig et al. 2012). The inverse transformation (a linear native space-to-midspace transform combined with a nonlinear midspace-to-MNI-space warp) was used to register all tractography results into a common template space for viewing.

Tract-based spatial statistics

DTI is a structural neuroimaging method that is sensitive to the diffusion of water. Because water diffusion is more directionally constrained and homogenous inside than outside an axon bundle, DTI measurements can be used to trace white matter pathways and to identify changes in those pathways over time (for a primer, see Mori and Zhang, 2006). FSL's Tract-Based Spatial Statistics (TBSS) package (Smith et al. 2006) was used to identify regions of white matter with changes in white matter integrity over time. This algorithm has been successfully used by multiple previous studies to identify changes in white matter and links between white matter variation and behavioral variation (e.g., Scholz et al. 2009; Schlegel et al. 2012; Lee et al. 2010; Taubert et al. 2010). TBSS performs voxelwise statistical analysis of fractional anisotropy (FA) images. FA reflects constraints on the directionality of water diffusion and is thought to be related to the degree of axon myelination, axon diameter, and/or tract density (Zatorre et al. 2012). Each subject's FA images for each time point were first nonlinearly registered to the FMRIB 1 mm^3 FA template, which is an MNI-space average of FA images from 58 subjects. All subjects' template-aligned FA images

were then averaged to create a study-specific mean FA image at all time points. This mean FA image was thinned to create a mean FA skeleton representing the centers of all white matter tracts common to the group. Aligned FA data from each subject at each time point was then projected onto this skeleton for voxelwise cross-time point statistical comparisons.

Previous functional studies of stone toolmaking found task-related activations in inferior frontoparietal regions (Stout and Chaminade 2007; Stout et al. 2008, 2011), so statistical comparison was constrained to the portion of the white matter skeleton that fell within the white matter beneath lateral frontoparietal cortex. This ROI was based on the superior longitudinal fasciculus (SLF) ROI from the JHU White Matter Atlas included in FSL, which was generated using probabilistic tractography in 28 subjects (Wakana et al. 2007; Hua et al. 2008). While the atlas's probabilistic SLF ROI terminates around the vicinity of premotor cortex, previous functional imaging studies have indicated that more anterior regions of inferior frontal gyrus within the atlas may be involved in toolmaking (Stout et al. 2008, 2011; Stout and Chaminade 2012). Monkey tract-tracing studies have shown that these anterior ventrolateral prefrontal regions do indeed receive white matter connections from the SLF (Schmahmann et al. 2007; Petrides and Pandya 2009), and diffusion imaging studies have successfully tracked these connections in both monkeys and humans (Schmahmann et al. 2007; Hecht et al. 2013a; Thiebaut de Schotten et al. 2012). Therefore, we thresholded the atlas's SLF ROI at 25 % and then manually extended it toward the anterior aspect of the inferior frontal gyrus. The threshold of 25 % was chosen as a conservative compromise between including voxels belonging to any subject in the probabilistic atlas (which would reflect a wide range of inter-subject variability) and including only voxels common to all subjects in the probabilistic atlas (which would create a very small ROI). This ROI was then used to mask the mean FA skeleton. As a control tract, we included the forceps major ROI from the JHU White Matter Atlas (Wakana et al. 2007; Hua et al. 2008), also thresholded at 25 %. Like the SLF, the forceps major is a cortico-cortical tract. It carries callosal connections between the left and right primary visual cortices and other early visual areas. To our knowledge there is no reason to suspect that this tract would be altered by toolmaking training. A single analysis was carried out on one mask that included both the SLF and forceps major in order to use a single test to establish whether toolmaking training had an effect on the SLF but not the forceps major (Nieuwenhuis et al. 2011).

Statistical comparisons were carried out using *randomize*, FSL's tool for permutation-based testing with a general linear model (GLM) (Nichols and Holmes 2002; see

Friston et al. 1994 for a discussion of the integration of general linear models with statistical parametric maps in neuroimaging). GLMs were constructed to compare FA changes along the 3 time points, from Time 1 to Time 2, Time 2 to Time 3, and Time 1 to Time 3 (i.e., not considering Time 2), both across the whole brain and in the SLF ROI. The GLMs incorporated the time point of each scan for each subject in days, with Time 1 being day 0. Threshold-free cluster enhancement (TFCE) (Smith and Nichols 2009) was used to identify statistically significant clusters of voxels. Results were corrected for multiple comparisons and thresholded at $p < .05$.

In clusters that showed significant FA change over time, we also examined axial and radial diffusivity. Fractional anisotropy is a measure of diffusivity in the primary diffusion direction (axial diffusivity) relative to diffusion in the other two perpendicular directions (radial diffusivity: the mean of these two latter measurements); changes in FA can therefore be due to underlying changes in axial and/or radial diffusivity. Preliminary research suggests that these may correspond to different biological mechanisms of change, although additional research in this area is warranted (Wheeler-Kingshott and Cercignani 2009). Increased axial diffusivity may reflect increases in axon density, axon caliber, microtubule packing, and/or microtubule organization (Kumar et al. 2010, 2012; Song et al. 2002, 2003; Choe et al. 2012), whereas decreased radial diffusivity may reflect increased axon myelination (Keller and Just 2009; Zhang et al. 2009; Bennett et al. 2010).

Radial diffusion images were created by averaging the images representing the magnitude of diffusion in the second and third directions; axial diffusion images were generated automatically by FSL in the standard DTI preprocessing pipeline. Both radial and axial images were then subjected to the same registration and processing steps as the FA images in order to enable measurement in equivalent voxels within and across subjects.

Voxel-based morphometry

Voxel-based morphometry is an approach to identifying structural differences in gray matter between subjects or within subjects across time which involves registering all scans to a common space and then measuring the spatial difference between each scan and the group average (for a primer, see Ashburner and Friston 2000). FSL's VBM tool (Douaud et al. 2007) was used to compare gray matter volume across time points. This algorithm performs voxelwise statistical analysis of T1-weighted MRI images. Scans were segmented into CSF, white matter, and gray matter, with the latter voxels being retained for further analyses. A study-specific, bilaterally symmetric gray matter template was produced by nonlinearly registering

each scan to the MNI T1 template, flipping each scan along the midline, and then averaging all scans. Each subject's native gray matter data was then nonlinearly registered to this template and modulated to correct for local expansion or contraction resulting from nonlinear registration, and then smoothed using a Gaussian kernel with a sigma of 3 mm (which is equivalent to FWHM approximately 7.08 mm). VBM compares individual subjects' template-space images to the template-space average, and intensity differences between the two indicate differences in gray matter volume. Analysis was constrained to gray matter regions connected by SLF (BA 45, 44, 6, 40, 39, and 37—the regions of experimental interest) and regions of early visual cortex (BA 17 and 18), since these regions are linked by the forceps major, the tract used as a control ROI in the TBSS analysis. As in the TBSS analysis, both the experimental and control regions were combined into one mask and subjected to a single statistical test in order to establish whether toolmaking training effected the gray matter areas connected by the SLF but not the gray matter areas connected by the forceps major (Nieuwenhuis et al. 2011). Voxelwise statistical comparisons of gray matter expansion/contraction were carried out using the same GLMs as the TBSS analyses, and threshold-free cluster enhancement (TFCE) (Smith and Nichols 2009) was used to identify statistically significant clusters of voxels. Clusters were thresholded at $p < .01$ but were not corrected for multiple comparisons.

Tractography from regions of white matter with significant change in fractional anisotropy

In order to obtain information about the anatomical connectivity of clusters showing significant changes in fractional anisotropy in the TBSS analysis, we used these clusters as seeds for further tractography analyses. Tractography was performed on each seed individually. Tracts were thresholded, normalized, binarized, and summed to produce composite images following the same approach as in the control tractography analyses. These tracts were produced to give anatomical elaboration to the TBSS results, and were not statistically compared across time points since they were the result of a previous longitudinal statistical analysis.

Tractography was performed in native diffusion space using FSL's *BEDPOSTX* and *probtrackx* tools (Behrens et al. 2003). *BEDPOSTX* uses Markov chain Monte Carlo sampling of the raw diffusion information to build a Bayesian distribution diffusion information at each voxel. *Probtrackx* samples from these distributions in order to generate probabilistic streamlines; in the resultant tractography output images, the intensity at each voxel corresponds to the number of streamlines that reached that voxel

(so greater intensity corresponds to a greater probability of connectivity within the model). Both *BEDPOSTX* and *probtrackx* model diffusion information not only along the principal direction of diffusion, but also along the other two orthogonal axes. Importantly, sampling of the Bayesian distributions for these additional diffusion directions allows tractography through regions with low fractional anisotropy, i.e., in regions of crossing fibers within white matter, and through the gray matter/white matter boundary (Behrens et al. 2007).

Results

Changes in white matter fractional anisotropy

FA changes occurred in white matter leading into left supramarginal gyrus (SMG), bilateral ventral precentral gyrus (vPrCG), and right inferior frontal gyrus *pars triangularis* (IFGpt). FA increased in these locations from Time 1 to Time 2 and decreased in almost identical locations from Time 2 to Time 3, although note that the vPrCG increase occurred in the right hemisphere and the vPrCG decrease occurred in almost exactly the same location in the left hemisphere (Fig. 2). No changes were observed in the forceps major. Coordinates, size, baseline FA values, and t values for FA changes for each cluster are reported in Table 2. The connectivity of each of these clusters is shown in Fig. 8 (described below).

In the clusters that showed significant FA change, we also measured axial and radial diffusivity in order to determine whether change in one or both of these measures was driving change in FA. From Time 1 to Time 2, axial diffusivity showed a significant increase (paired samples t test; $t(17) = 5.299$, $p < .001$, 2-tailed), while radial diffusivity showed no significant change [paired samples t test; $t(17) = -.954$, $p = .353$, 2-tailed]. From Time 2 to Time 3, axial diffusivity showed a significant decrease [paired samples t test; $t(17) = -2.565$, $p = .020$, 2-tailed], while radial diffusivity again showed no significant change [paired samples t test; $t(17) = -.596$, $p = .559$, 2-tailed], suggesting that FA changes may have been due to changes in axon density, axon caliber, microtubule packing, and/or microtubule organization, rather than changes in myelination (Keller and Just 2009; Zhang et al. 2009; Bennett et al. 2010; Kumar et al. 2010, 2012; Song et al. 2002, 2003; Choe et al. 2012).

Correlation with training time

White and gray matter changes mirrored the intensive training (mean 120 h) from Time 1 to Time 2 and the relative decrease (mean 47 h) from Time 2 to Time 3 (see

Fig. 1). Total FA change (ΔFA) was correlated with training hours since the previous scan ($r = .580$, $p = .024$, 1-tailed) (Fig. 3). ΔFA values within each cluster individually were not significantly correlated with training time, likely reflecting individual variability in the allocation of structural responses (Sect. "Variability in FA change across subjects").

Correlation with motor performance

Across all time points, accuracy ranged from .74 to .96 (mean = .86, SD = .05). Accuracy at Time 3 did not differ significantly from Time 1 (paired t test, $t = 1.717$, $p = .161$), reflecting substantial variance in individual differences (range +.10 to −.04). Total FA change (ΔFA) was correlated with changes in accuracy (Δaccuracy $r = .616$, $p = .029$, 1-tailed) (Fig. 4). Performance differed ($t = 3.226$, $df = 8$, $p = .012$) between the two subject groups identified from PCA factor scores, with Group 1 showing a mean Δaccuracy of +4 % across time points and Group 2 a mean of −1 %. Thus, Group 1 subjects improvement their striking accuracy over the training period (mean = +8 %, paired $t = 5.869$, $df = 2$, $p = .028$), whereas Group 2 subjects did not (mean = −1 %, paired $t = -.467$, $df = 1$, $p = .722$). From Time 1 to Time 2, Δaccuracy correlated with PCA Component 2 factor scores ($r = .878$, $p = .025$, 1-tailed), whereas from Time 2 to Time 3 it correlated with Component 1 ($r = .973$, $p = .003$, 1-tailed). Analysis at the level of individual clusters is consistent with this: from Time 1 to Time 2 Δaccuracy is positively correlated with right vPrCG ($r = .822$, $p = .044$, 1-tailed) and negatively correlated with right IFGpt ($r = -.989$, $p = .001$, 1-tailed), whereas from Time 2 to Time 3, Δaccuracy is positively correlated with right IFGpt ($r = .837$, $p = .039$, 1-tailed).

Variability in FA change across subjects

Closer examination of each subject's pattern of FA changes across clusters reveals substantial individual variability (Fig. 5). A principal components analysis identified 2 components accounting for 83 % of this variance (Fig. 6). Component 1 (61 % of variance) captures the covariance of all clusters apart from the right vPrCG, while component 2 (22 % of variance) primarily captures variation in right vPrCG, including a negative correlation with right IFGpt. Inspection of individual factor scores for component two suggests the presence of two alternative responses to training. Group 1 (subjects 1,3, 5 and 6) is characterized by positive loadings from Time 1 to Time 2 and Group 2 (subjects 2 and 4) by negative loadings (Fig. 7). For all subjects, this pattern reverses from Time 2 to Time 3, reflecting the symmetric decreases in FA noted above
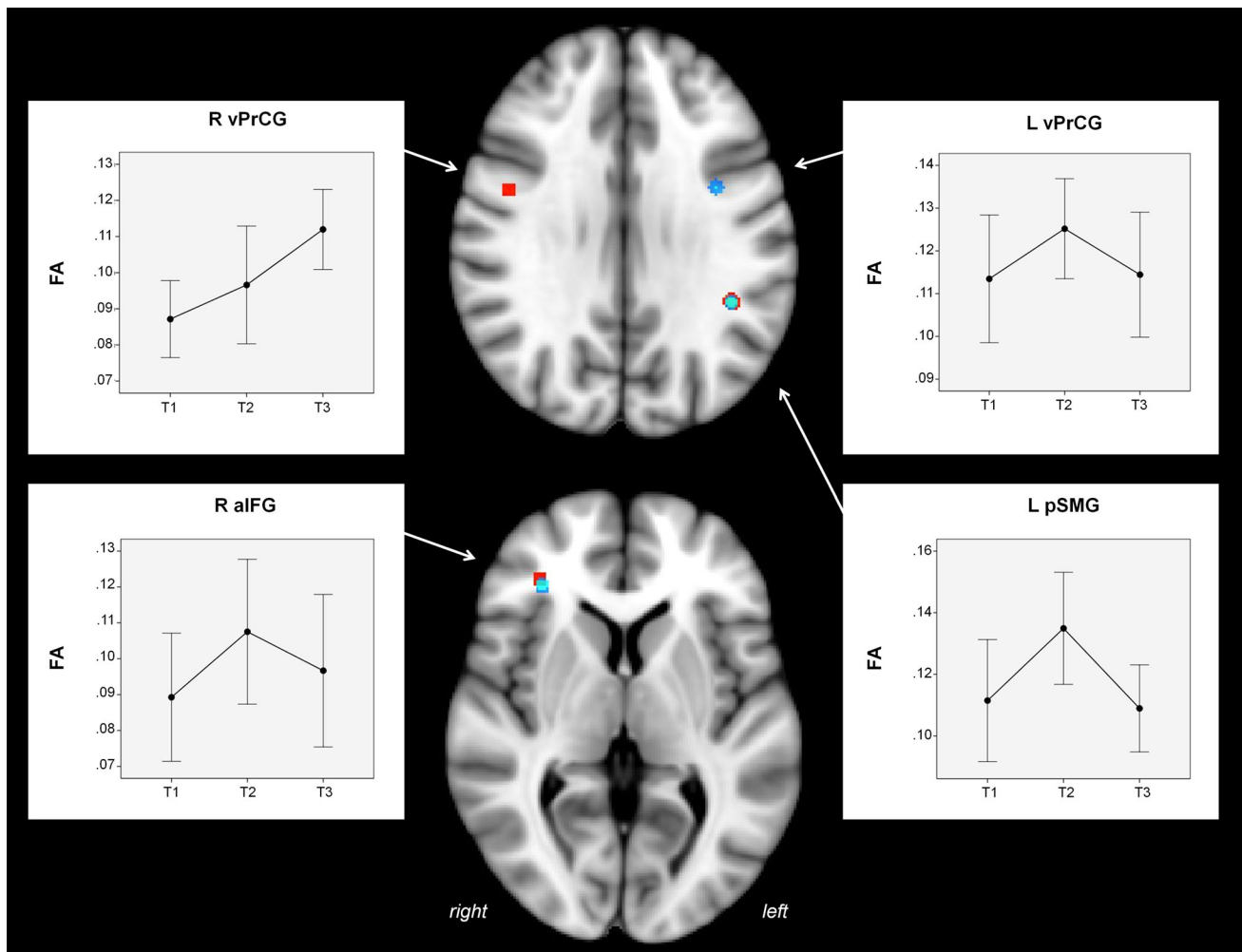
Fig. 2 Structural changes associated with the acquisition of Paleolithic stone toolmaking skills. Changes in SLF *white* matter ($p < .05$, corrected). *Yellow–red* increases from T1 to T2; *cyan–blue* decreases from T2 to T3. *Insets* show mean FA at each time point. Note the close overlap in the spatial location of increases/decreases in IFGpt and SMG; increases/decreases were observed in very similar locations of vPrCG in each hemisphere. While FA within the R vPrCG cluster continued to increase from T2 to T3, note that this was not significant (i.e., the voxelwise analysis that identified clusters showing significant change from T2 to T3 did not find such an increase at this location). *R/L vPrCG* right/left ventral precentral gyrus, *R aIFG* right anterior inferior frontal gyrus, *L pSMG* left posterior supramarginal gyrus

(Fig. 2). Training time was not correlated with PCA factor scores.

Changes in gray matter volume

The VBM analysis in regions surrounding the SLF found gray matter expansion from Time 1 to Time 2 in the supramarginal gyrus, and a reversal of this effect from Time 2 to Time 3 (Fig. 8c), mirroring the pattern of change in the TBSS analysis. This effect was significant at the $p < .01$ level without correction for multiple comparisons, and should therefore be considered a trend. At this threshold there were no voxels showing gray matter change over time in visual cortex.

Probabilistic tractography from regions of white matter change

Probabilistic tractography confirmed that the voxels showing white matter change projected to the voxels showing gray matter change. Cortical regions identified by tractography and morphometry are known to be functionally responsive to stone toolmaking (Stout and Chaminade 2007; Stout et al. 2008, 2011) (Fig. 8).

Table 2 Baseline FA, t scores for FA change, volume, and coordinates for clusters showing significant change in FA over time

	Baseline FA at time 1	Max t score in cluster for change in FA	Volume (mm³)	Coordinates of peak: voxels (mm)		
				x	y	z
Clusters from Time 1–Time 2 analysis						
L pSMG	.141127	17.5	73	132 (−42)	83 (−43)	103 (31)
R aIFG	.184215	18.5	35	56 (34)	164 (38)	77 (5)
R vPrCG	.142852	13.1	25	44 (46)	127 (1)	102 (30)
Clusters from Time 2–Time 3 analysis						
L pSMG	.142939	14.7	93	132 (−42)	83 (−43)	103 (31)
R aIFG	.139591	13.3	44	57 (33)	162 (36)	78 (6)
L vPrCG	.158409	15.8	147	126 (−36)	128 (2)	101 (29)

Coordinates are given in 1 mm MNI template space
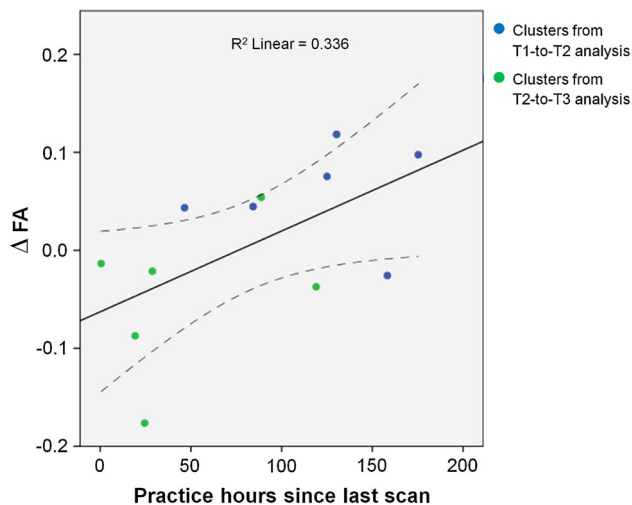
Fig. 3 Correlation between change in FA and practice hours. *Gray lines* denote 95 % confidence interval. Pearson's $r = .580$, $p = .024$, 1-tailed
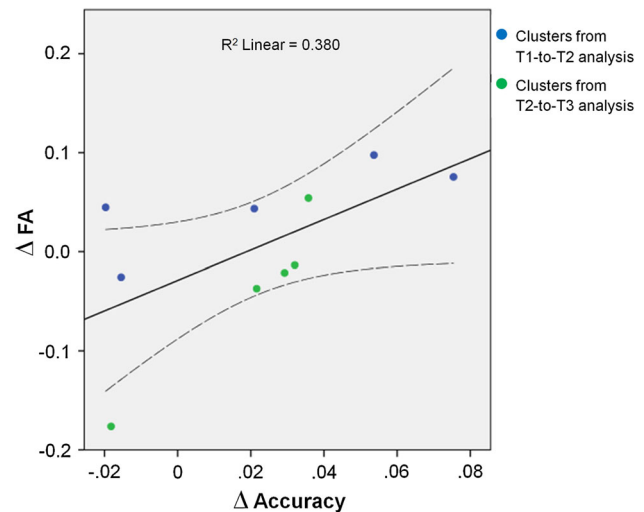
Fig. 4 Correlation between change in FA and change in accuracy on striking task. *Gray lines* denote 95 % confidence interval. Pearson's $r = .616$, $p = .029$, 1-tailed

Discussion

Skill acquisition and structural remodeling

These results establish that the acquisition of Stone Age toolmaking skills causes structural remodeling to fronto-parietal circuits. This is in agreement with previous studies that have found structural changes with skill acquisition [e.g., (Maguire et al. 2000; Draganski et al. 2004; Floyer-Lea and Matthews 2004; Lappe et al. 2008; Scholz et al. 2009; Lee et al. 2010; Taubert et al. 2010, 2011; Sisti et al. 2012; Steele et al. 2012)]. The robustness of the anatomical changes we observed is underscored by the fact that they reached statistical significance in a relatively small sample size. This is likely related to the intensity and duration of our training program. Previous studies have reported significant results from larger samples using considerably shorter, more limited, and less evolutionarily relevant training regimes. For example, in a group of studies on

balancing (Taubert et al. 2010, 2011, 2012), 14 subjects received six 45-min training sessions over the course of 6 weeks. In another study (Steele et al. 2012), 13 subjects were trained on a motor sequence task over a period of only 5 days. Significant effects have also been obtained using relatively small samples (<12) (Floel et al. 2009; Takeuchi et al. 2010; Antonenko et al. 2012; Gebauer et al. 2012; Schlegel et al. 2012; Wang et al. 2012; Tseng et al. 2013). In our study training was both intensive and long-term, extending over a period of 2 years. This enabled us to study the complex, real-world skill of stone toolmaking, which can take years to master (Stout 2002). The current study thus applied a well-established experimental approach to address for the first time questions about human brain evolution. We did so by studying an evolutionarily relevant skill that can be archeologically tied to particular times and places in prehistory.

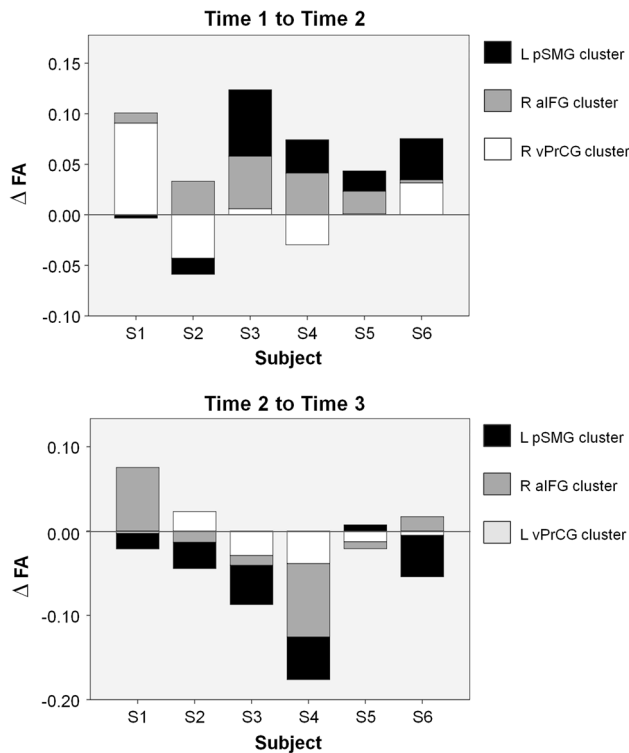The cellular mechanisms responsible for experience-dependent gray and white matter change are under active

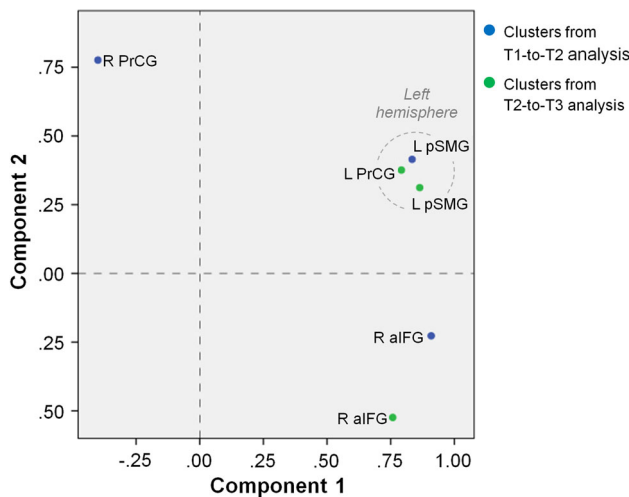Fig. 5 Individual variation in pattern of FA change across clusters

Fig. 6 Factor loadings of each cluster in each of two components identified by principal components analysis. Note that the left hemisphere clusters have very similar factor loadings (*dashed circle*)

Fig. 7 Individual subjects' factor scores for Component 2. Note that subjects are arranged on the *x*-axis to emphasize the distinction between Subjects 2 and 4 versus the others

training. Training time between scans varied across subjects and predicted total ΔFA. This included increases in total FA associated with intense training from Time 1 to Time 2 as well as symmetrical decreases in FA associated with lower levels of training from Time 2 to Time 3. A trend toward gray matter expansion and contraction observed in the supramarginal gyrus followed the same pattern. These sensitive anatomical responses likely reflect the acute nature of behavioral demands on brain structure, a finding which is supported by a growing body of evidence: white matter changes can occur in as little as 2 h of behavioral training in humans (Sagi et al. 2012), and changes in fractional anisotropy of white matter can be detected in as little as 1 h in mice (Ding et al. 2013). The fact that white and gray matter changes rapidly reversed with reduction of toolmaking activity may also reflect the metabolic costliness of maintaining such changes. This acute response confirms the structurally demanding nature of Paleolithic toolmaking behavior, even in our modern human subjects, and is consistent with toolmaking's hypothesized role in generating selective pressure on frontoparietal cortex (see below).

Total ΔFA also correlated with changes in striking accuracy across subjects. Striking accuracy is essential in stone toolmaking to control fracture and achieve desired effects. Experimental studies of modern stone craftsmen have shown that control over the elementary striking gesture is a more reliable indicator of expertise than years of experience or amount of technical knowledge (Bril et al. 2005). Stone toolmakers must adapt motor performance to variable conditions (e.g., composition, size, and shape of both percussor and core), and experimental manipulations

investigation, but are not yet well understood (Zatorre et al. 2012). Recent reviews (e.g., Thomas and Baker 2013) highlight the need to ensure that structural change over time is due to training and not some other temporally varying process such as exam schedules or equipment change. In the current study, structural changes correlated with both practice hours and motor performance, confirming their causal association with stone toolmaking
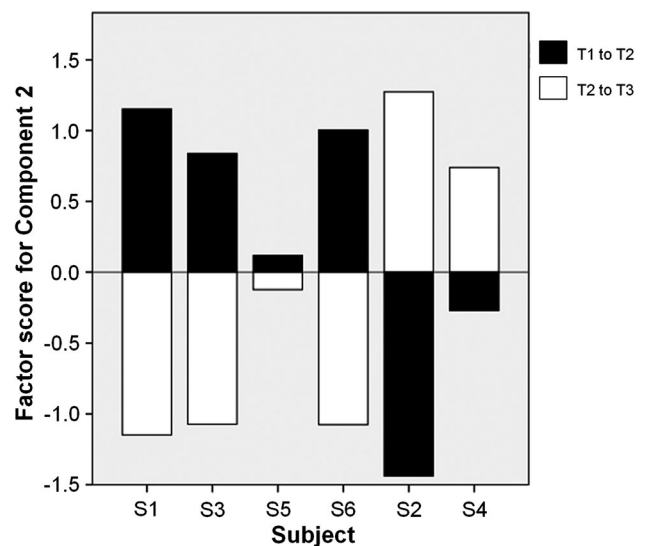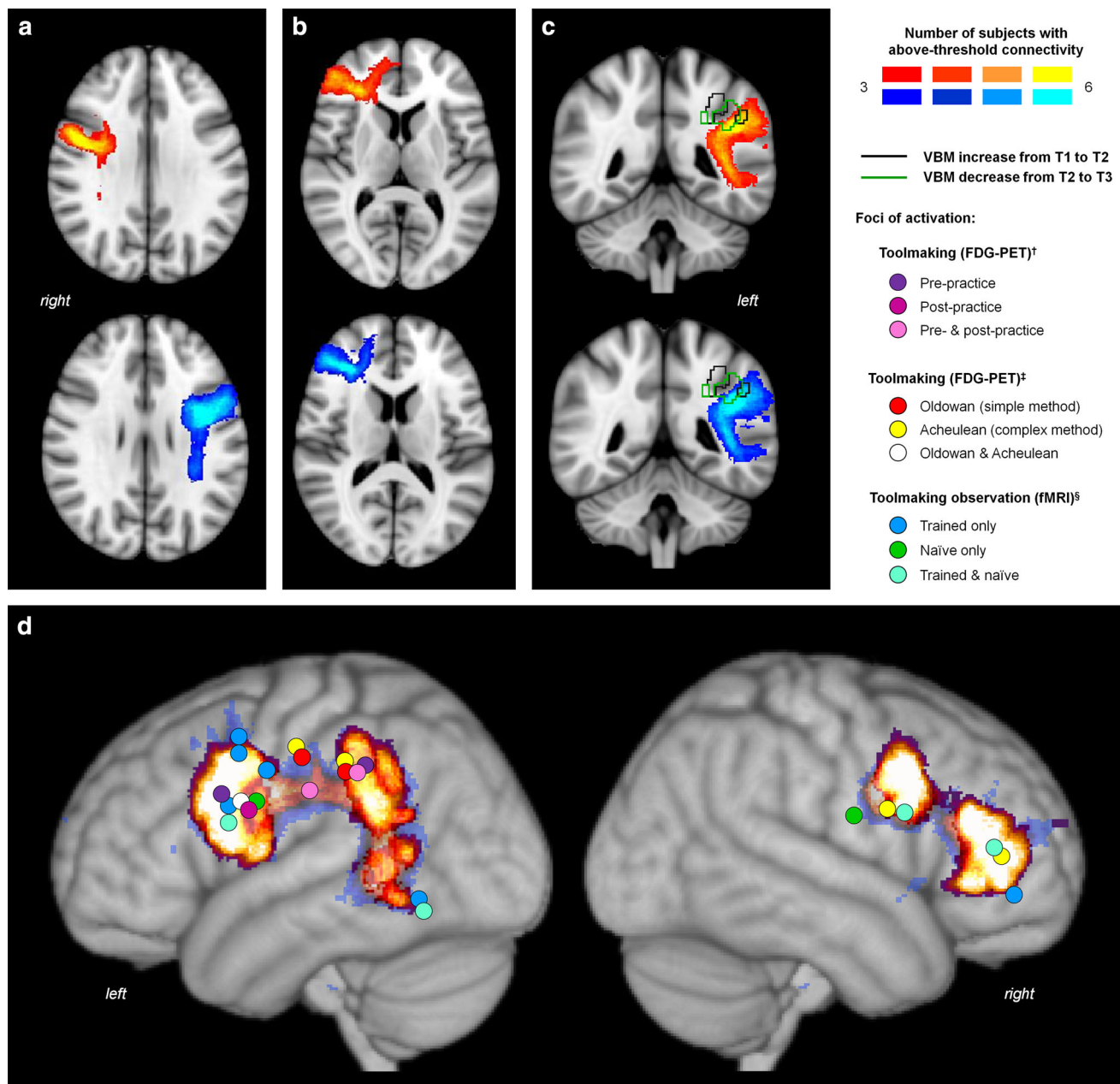
Fig. 8 Regions of *white* matter change project to regions of gray matter change, and to regions of *gray* matter activated in previous Paleolithic toolmaking studies. Clusters of white matter change (Fig. 2) were used to seed tractography analyses. In **a–c**, hot scale indicates increases from T1 to T2; cool scale indicates decreases from T2 to T3. **d** Shows the same tracts, all made partially transparent and overlaid. **a** Tracts produced after seeding with the L and R vPrCG TBSS clusters. These projections connect the body of the SLF to the cortex of the ventral precentral gyrus. **b** Tracts produced after seeding with the R aIFG clusters. These projections connect the pars triangularis of the right IFG with the right dorsomedial frontal pole. **c** Tracts produced after seeding with the L pSMG clusters. These projections connect the posterior inferior parietal cortex with lateral temporal cortex and project into a region of gray matter which showed a trend toward VBM increases from T1 to T2 and decreases from T2 to T3 ($p < .01$, uncorrected). **d** All tractography results rendered on the MNI template brain. *Dots* indicate activations related to Paleolithic toolmaking from previous studies that fall within or immediately adjacent to tractography results from the present study. [†] Stout and Chaminade (2007); [‡]Stout et al. (2008); [§]Stout et al. (2011)

have documented the flexible adaptation of skilled tool-makers (Bril et al. 2005, 2010). We examined the transfer of motor skill from toolmaking training to a controlled virtual reality striking task and found that individual changes in accuracy (rate of hitting the target) correlated

with FA change. In a naturalistic study, Nonaka et al. (2010) had subjects strike stone cores to remove flakes and found that the mean distance between predicted and actual points of impact for novice toolmakers was 7.4 mm, with a SD of approximately 4.6 mm (see Fig. 6 in Nonaka et al.

2010). Assuming a normal distribution, this implies that approximately 84 % of strikes would have been within 12 mm of the target (mean +1 SD), which is comparable to the mean hit rate in our experiment (86 %) using targets with an average diameter of 15 mm. Our most successful subject achieved a hit rate of 96 % by Time 3, which is roughly that expected for the "intermediate" subjects of Nonaka et al. (mean error = 4.3 mm, SD = 5.7, mean +2 SD = 15.7), but falls below the expected rate for "experts" (mean error = .6 mm, SD = 1.1, mean + 13SD = 14.9) which should approximate 100 %. To the extent that increases in FA are correlated with increases in accuracy, this suggests that the training-related anatomical adaptations we observed may be relatively modest compared to those that would have been experienced by expert Pleistocene toolmakers (compare Fig. 1b vs. c, see also Stout et al. 2014).

In contrast to total ΔFA, training time did not correlate with ΔFA in individual clusters. This reflects individual variation in the specific response to training. Despite the limitations of our small sample size for investigating individual variation, we did observe apparent patterning that might be further investigated in future research. TBSS analysis identified six clusters of significant FA change at the group level, but also substantial individual variability in the relative allocation of structural changes across these clusters. A Principal Components Analysis revealed two main dimensions of this variation, including a pattern of coordinated change across all clusters apart from right vPrCG (Component 1), and an inverse relationship between right vPrCG and right IFGpt (Component 2). Changes in striking accuracy were predicted by Component 2 from Time 1 to Time 2 and by Component 1 from Time 2 to Time 3. In other words, initial increases in striking accuracy were associated with increased right vPrCG FA at the expense of right IFGpt whereas subsequent increases in accuracy were associated with retention of FA increases (i.e., less negative ΔFA) outside right vPrCG, especially in right IFGpt. Changes in accuracy did not correlate with training time, in keeping with the disjunction between stone working skill and experience reported by Bril et al. (2005) and again reflecting a variable response to training across our subjects.

Interestingly, subjects fell into two groups with respect to Component 2 scores. Group 1 showed positive scores from Time 1 to Time 2 followed by negative scores from Time 2 to Time 3 whereas Group 2 showed the opposite pattern. Group 1 subjects improvement their striking accuracy over the training period whereas Group 2 subjects did not, suggesting that the former pursued a more effective learning strategy associated with a different pattern of structural brain changes. However, our data do not allow us to test the possibility that Group 2 subjects prioritized other aspects of skill learning (e.g., conceptual knowledge) that we were unable to measure, rather than simply displaying less effective learning overall.

This variable pattern is consistent with the known involvement of right inferior frontal cortex in action programming, and with the context-sensitive nature of interactions between IFG, the ventral premotor cortex or vPrCG, and primary motor cortex. Whereas ventral premotor cortex directly modulates primary motor activity, right IFGpt is thought to play a role in the more abstract cognitive control of action [e.g., inhibition under conditions of response uncertainty (Levy and Wagner 2011)] through its extensive projections to ventral premotor cortex (Dum and Strick 2005). Using a paired-pulse transcranial magnetic stimulation paradigm, Buch et al. (2010) found that ventral premotor cortex variably facilitated or inhibited primary motor corticospinal activity, depending on task context (reach vs. switch targets), and that individual variation in the magnitude of these effects correlated with FA in right precentral and inferior frontal gyrus clusters closely approximating (<9.2 mm Euclidean distance) those identified here. Our results suggest that anatomical changes in right vPrCG and IFGpt may play somewhat different roles in the acquisition of complex skills, perhaps corresponding to early vs. late learning processes and/or different learning objectives (e.g., motor skill vs. conceptual understanding).

Toolmaking, neural plasticity, and brain evolution

Previous functional imaging research (Stout and Chaminade 2007; Stout et al. 2008, 2011) has established that Paleolithic toolmaking by modern human subjects recruits portions of frontoparietal cortex which have been identified by comparative studies (Peeters et al. 2009, 2013; Orban and Rizzolatti, 2012; Hecht et al. 2013a, b; Rilling et al. 2008; Ramayya et al. 2010) as likely sites of human-specific adaptations for tool use. Current results demonstrate that active practice of Paleolithic toolmaking skills elicits structural remodeling in these same regions (Figs. 2, 8), providing further confirmation for hypothesized links between stone toolmaking and frontoparietal brain structure and function. The observation of anatomical effects in our subjects indicates that, even in large-brained, modern humans, stone toolmaking is sufficiently demanding to induce the metabolically costly re-allocation of structural resources in ways that parallel inferred evolutionary changes (increased frontoparietal gray matter volume and connectivity). Habitual toolmaking by smaller-brained Paleolithic ancestors would have stressed these anatomical substrates to a similar or even greater degree, generating positive selection pressure on any genetic variants that enhanced the reliability and efficiency of the plastic phenotypic response.

This potential role of phenotypic plasticity in facilitating evolutionary adaptation has been appreciated since the nineteenth century (Baldwin 1896; Osborn 1896) and is now widely known as the "Baldwin effect" (Weber and Depew 2003; Bateson 2004). Briefly, phenotypic plasticity allows organisms to adapt to new environmental conditions and/or generate novel adaptive behaviors. In the context of these new conditions or behaviors, any variation in the ease, efficiency, or reliability of the expression of the plastic trait will be acted on by selection. This would potentially lead to alterations in the genes regulating the trait and to the inheritance of a predisposition to develop the modified condition more or less independently of the original environmental/behavioral stimulus. We suggest that such a process may have occurred over the 2.5 million years during which stone toolmaking was a key adaptive skill for our ancestors.

The earliest, Oldowan, stone tools consist of sharp stone flakes struck from river cobbles (Semaw et al. 2003). This early technology predates fossil evidence of substantial hominin brain expansion, although it has been suggested that size-independent adaptations (i.e., "re-organization") of posterior parietal and inferior frontal cortex may already have been present (Holloway et al. 2004a, b). Functional investigations of Oldowan toolmaking (Stout and Chaminade 2007; Stout et al. 2008) observed activations in vPrCG and in SMG which are now closely paralleled by evidence of training-related structural white and gray matter changes reported here. Fossil, functional and structural evidence are thus consistent with the hypothesis (Stout and Chaminade 2007) that early hominin toolmaking was supported by evolutionary elaborations of a primitive ventral frontoparietal circuit for object manipulation that is shared with other primates (Maravita and Iriki 2004; Rizzolatti et al. 1998). Indeed, a VBM study of tool-use (rake) learning in macaques found evidence of parietal gray matter expansion encompassing potions of anterior intraparietal sulcus (Quallo et al. 2009) that may be comparable to the human gray matter expansion reported here. It has previously been suggested (Iriki and Taoka 2012) that such plasticity, already present in monkeys, may have contributed to the evolution of human inferior parietal areas through the Baldwin effect. Our results extend this argument to include frontoparietal white matter connections, and link it to specific, archeologically observable, behaviors known to have occurred during human evolution.

In addition to changes in white matter underlying ventral PrCG and SMG, we observed a significant effect under right IFGpt. Previous functional studies have shown increased activation of this region by Acheulean, but not Oldowan, toolmaking (Stout et al. 2008, 2011). Early Acheulean toolmaking first appeared around 1.75 million years ago (Lepre et al. 2011; Beyene et al. 2013), roughly .25 million years after the first widely accepted fossil evidence of brain expansion and inferior frontal re-organization (Holloway et al. 2004a, b). The Acheulean is widely considered to represent a major advance in toolmaking sophistication over the preceding Oldowan, providing the first clear evidence of artifacts ("handaxes", "picks") made to a predetermined design. Acheulean tools also show evidence of increasing refinement through time (Beyene et al. 2013) and by 500,000 years ago (Fig. 1c) their production involved an elaborate sequence of hierarchically embedded goals and sub-goals (Stout 2011; Stout et al. 2014). This later Acheulean time period (780–400,000 years ago) coincides with the fastest rate of hominin encephalization in the past 2 million years (Ruff et al. 1997), and includes the form of "Late Acheulean" toolmaking in which our research subjects were trained (Fig. 1b). In keeping with evidence that the basic manipulative complexity of Oldowan vs. Late Acheulean toolmaking does not differ (Faisal et al. 2010), we have previously interpreted (Stout et al. 2008, 2011) increased right inferior frontal gyrus pars triangularis (IFGpt) activation as reflecting the cognitive control [e.g., inhibition, task switching (Aron et al. 2004; Koechlin and Jubault 2006; Levy and Wagner 2011)] demands of the more complex action sequences involved in this technology. Current evidence of training-related FA changes in white matter under right IFGpt corroborates functional evidence of this region's involvement in stone toolmaking, and provides the first direct evidence that Paleolithic toolmaking can elicit measurable structural responses in prefrontal cortex.

Toolmaking and language

The frontoparietal regions in which we observed structural change participate not only in tool use, but also in other complex cognitive tasks including action planning and language. Loci of overlap between praxis and communication have been identified in left ventral premotor (Rizzolatti and Arbib 1998) and parietal (Frey 2008) cortex, and both are now specifically linked to Paleolithic toolmaking by functional (Stout and Chaminade 2007; Stout et al. 2008, 2011) and structural results. This supports the hypothesized co-evolution of tool use and other complex functions (Rizzolatti and Arbib 1998; Corballis 2002; Pulvermüller and Fadiga 2010; Stout and Chaminade 2012). Our results also corroborate functional evidence of right IFGpt involvement in toolmaking (Stout et al. 2008, 2011). Right IFGpt is not a core language or tool use region, but supports domain-general cognitive control (Vigneau et al. 2011) relevant to a range of complex communicative [e.g., discourse comprehension (Menenti et al. 2009)] and instrumental [e.g., task switching (Aron

et al. 2004; Koechlin and Jubault 2006; Levy and Wagner 2011)] behaviors.

Structural remodeling to these regions in response to Paleolithic toolmaking is consistent with longstanding models of the mutually reinforcing interaction between technological, social, communicative, and neural complexity in human evolution (Holloway 1967; Engels 2003 [1876]). More specifically, we propose that human frontoparietal circuits underwent adaptations for Paleolithic toolmaking that were behaviorally co-opted ["exapted" (Gould and Vrba 1982)] to support proto-linguistic communication and then subsequently altered by secondary adaptations specific to language.

References

Andersson JLR, Jenkinson M et al (2007). Non-linear optimisation: FMRIB technical report TR07JA1 from www.fmrib.ox.ac.uk/analysis/techrep

Antonenko D, Meinzer M et al (2012) Grammar learning in older adults is linked to white matter microstructure and functional connectivity. Neuroimage 62(3):1667–1674

Arbib MA (2005) From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. Behav Brain Sci 28(2):105–124

Aron AA, Robbins TW et al (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8(4):170–177

Ashburner J, Friston KJ (2000) Voxel-based morphometry—the methods. Neuroimage 11(6 Pt 1):805–821

Avants BB, Schoenemann PT, James CG (2006) Lagrangian frame diffeomorphic image registration: morphometric comparison of human and chimpanzee cortex. Med Image Anal 10(3): 397–412

Baldwin JM (1896) A new factor in evolution. Am Nat 30(354):441–451

Bateson P (2004) The active role of behaviour in evolution. Biol Philos 19(2): 283–298

Behrens TE, Woolrich MW et al (2003) Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 50(5):1077–1088

Behrens TE, Johansen-Berg H et al (2007) Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage 34(1):144–155

Bennett I, Madden DJ, Vaidya CJ, Howard DV, Howard JH Jr (2010) Age-related differences in multiple measures of white matter integrity: a diffusion tensor imaging study of healthy aging. Hum Brain Mapp 31:378–390

Beyene Y, Katoh S, WoldeGabriel G, Hart WK, Uto K, Sudo M, Kondo M, Hyodo M, Renne PR, Suwa G, Asfaw B (2013) The

characteristics and chronology of the earliest Acheulean at Konso, Ethiopia. PNAS 110(5):1584–1591

Bradley B, Sampson CG (1986) Analysis by replication of two Acheulian artefact assemblages. In: Bailey G, Callow P (eds) Stone Age Prehistory: Studies in Honour of Charles McBurney, pp 29–45

Bril B, Roux V, Dietrich G (2005). Stone knapping: Khambhat (India), a unique opportunity? Stone knapping: the necessary conditions for a uniquely hominin behaviour. In: Roux V, Bril B (eds) McDonald Institute for Archaeological Research, Cambridge, pp 53–72

Bril B, Rein R, Nonaka T, Wenban-Smith F, Dietrich G (2010) The role of expertise in tool use: skill differences in functional action adaptations to task constraints. J Exp Psychol Hum Percept Perform 36(4):825–839

Buch ER, Mars RB, Boorman ED, Rushworth MFS (2010) A network centered on ventral premotor cortex exerts both facilitatory and inhibitory control over primary motor cortex during action reprogramming. J Neurosci 30(4):1395–1401

Choe AS, Stepniewska I, Colvin DC, Ding Z, Anderson AW (2012) Validation of diffusion tensor MRI in the central nervous system using light microscopy: quantitative comparison of fiber properties. NMR Biomed 25:900–908

Corballis MC (2002) From hand to mouth: the origins of language. Princeton University Press, Princeton

Denys K, Vanduffel W et al (2004) Visual activation in prefrontal cortex is stronger in monkeys than in humans. J Cogn Neurosci 16(9):1505–1516

Ding AY, Li Q et al (2013) MR diffusion tensor imaging detects rapid microstructural changes in amygdala and hippocampus following fear conditioning in mice. PLoS One 8(1):e51704

Douaud G, Smith S et al (2007) Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. Brain 130(Pt 9):2375–2386

Draganski B, Gaser C et al (2004) Neuroplasticity: changes in grey matter induced by training. Nature 427(6972):311–312

Dum RP, Strick PL (2005) Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. J Neurosci 25(6):1375–1386

Engels F (2003 [1876]) The part played by labour in the transition from ape to man. In: Scharff RC, Dusek V (eds) Philosophy of technology. Blackwell, London, pp 71–77

Engvig A, Fjell AM et al (2012) Memory training impacts short-term changes in aging white matter: a longitudinal diffusion tensor imaging study. Hum Brain Mapp 33(10):2390–2406

Faisal AA, Wolpert DM (2009) Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. J Neurophysiol 101(4):1901

Faisal A, Stout D et al (2010) The manipulative complexity of Lower Paleolithic stone toolmaking. PLoS One 5(11):e13718

Falk D (1983) Cerebral cortices of East African early hominids. Science 221(4615):1072–1074

Floel A, de Vries MH et al (2009) White matter integrity in the vicinity of Broca's area predicts grammar learning success. Neuroimage 47(4):1974–1981

Floyer-Lea A, Matthews PM (2004) Changing brain networks for visuomotor control with increased movement automaticity. J Neurophysiol 92(4):2405–2412

Frey SH (2008) Tool use, communicative gesture and cerebral asymmetries in the modern human brain. Philos Trans R Soc B Biol Sci 363(1499):1951–1957

Friston KJ et al (1994) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2(4):189–210

Gebauer D, Fink A et al (2012) Differences in integrity of white matter and changes with training in spelling impaired children: a

diffusion tensor imaging study. Brain Struct Funct 217(3):747–760

Gould SJ, Vrba ES (1982) Exaptation—a missing term in the science of form. Paleobiology 8(1):4–15

Greenfield PM (1991) Language, tools, and brain: the development and evolution of hierarchically organized sequential behavior. Behav Brain Sci 14:531–595

Hecht EE, Gutman DA et al (2013a) Process versus product in social learning: comparative diffusion tensor imaging of neural systems for action execution-observation matching in macaques, chimpanzees, and humans. Cereb Cortex 23(5):1014–1024

Hecht EE, Murphy LE et al (2013b) Differences in neural activation for object-directed grasping in chimpanzees and humans. J Neurosci 33(35):14117–14134

Hill J, Inder T, Neil J, Dierker D, Harwell J, Van Essen D (2010) Similar patterns of cortical expansion during human development and evolution. Proc Natl Acad Sci 107(29):13135–13140

Holloway RL (1967) The evolution of the human brain: some notes toward a synthesis between neural structure and the evolution of complex behavior. Gen Syst 12:3–19

Holloway RL, Broadfield D, Yuan M (2004a) The human fossil record, vol 3, brain endocasts–the paleoneurological evidence. Wiley-Liss, Hoboken

Holloway RL, Broadfield D et al (2004b) The human fossil record, volume 3, brain endocasts–the paleoneurological evidence. Wiley-Liss, Hoboken

Hua K, Zhang J et al (2008) Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. Neuroimage 39(1):336–347

Iriki A, Taoka M (2012) Triadic (ecological, neural, cognitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. Philos Trans R Soc B Biol Sci 367(1585):10–23

Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5(2):143–156

Jenkinson M, Bannister P et al (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17(2):825–841

Jenkinson M, Beckmann CF et al (2012) FSL. Neuroimage 62(2):782–790

Keller TA, Just MA (2009) Altering cortical connectivity: remediation-induced changes in the white matter of poor readers. Neuron 64:624–631

Khreisheh NN, Davies D, Bradley BA (2013) Extending experimental control: the use of porcelain in flaked stone experimentation. Adv Archaeol Prac: J Soc Am Archaeol 1(1):37–46

Koechlin E, Jubault T (2006) Broca's area and the hierarchical organization of human behavior. Neuron 50(6):963–974

Kumar R, Macey PM, Woo MA, Harper RM (2010) Rostral brain axonal injury in congenital central hypoventilation syndrome. J Neurosci Res 88:2146–2154

Kumar R, Nguyen HD, Macey PM, Woo MA, Harper RM (2012) Regional brain axial and radial diffusivity changes during development. J Neurosci Res 90:346–355

Langergraber KE et al (2012) Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci 109(39):15716–15721

Lappe C, Herholz SC et al (2008) Cortical plasticity induced by short-term unimodal and multimodal musical training. J Neurosci 28(39):9632–9639

Lee B, Park JY et al (2010) White matter neuroplastic changes in long-term trained players of the game of Baduk (GO): a voxel-based diffusion-tensor imaging study. Neuroimage 52(1):9–19

Lepre CJ, Roche H, Kent DV, Harmand S, Quinn RL, Brugal J-P, Texier P-J, Lenoble A, Feibel CS (2011) An earlier origin for the Acheulian. Nature 477(7362):82–85

Levy BJ, Wagner AD (2011) Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. Ann N Y Acad Sci 1224(1):40–62

Maguire EA, Gadian DG et al (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci USA 97(8):4398–4403

Maravita A, Iriki A (2004) Tools for the body (schema). Trends Cogn Sci 8(2):79–86

Menenti L, Petersson KM et al (2009) When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. J Cogn Neurosci 21(12):2358–2368

Mori S, Zhang J (2006) Principles of diffusion tensor imaging and its applications to basic neuroscience research. Neuron 51(5):527–539

Nelissen K, Luppino G et al (2005) Observing others: multiple action representation in the frontal lobe. Science 310(5746):332–336

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15(1):1–25

Nieuwenhuis S, Forstmann BU, Wagenmakers EJ (2011) Erroneous analyses of interactions in neuroscience: a problem of significance. Nat Neurosci 14(9):1105–1107

Nonaka T, Bril B, Rein R (2010) How do stone knappers predict and control the outcome of flaking? Implications for understanding early stone tool technology. J Hum Evol 59(2):155–167

Oakley KPBM (1949) Man: the toolmaker. Trustees of the British Museum, London

Orban GA, Rizzolatti G (2012) An area specifically devoted to tool use in human left inferior parietal lobule. Behav Brain Sci 35(4):234

Osborn H (1896) A mode of evolution requiring neither natural selection nor the inheritance of acquired characters. Trans N Y Acad Sci 15:141–148

Peeters R, Simone L et al (2009) The representation of tool use in humans and monkeys: common and uniquely human features. J Neurosci 29(37):11523–11539

Peeters RR, Rizzolatti G et al (2013) Functional properties of the left parietal tool use region. NeuroImage 78:83–93

Petrides M, Pandya DN (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. PLoS Biol 7(8):e1000170

Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci 11(5):351–360

Quallo M, Price C, Ueno K, Asamizuya T, Cheng K, Lemon R, Iriki A (2009) Gray and white matter changes associated with tool-use learning in macaque monkeys. Proc Natl Acad Sci 106(43):18379–18384

Ramayya AG, Glasser MF et al (2010) A DTI investigation of neural substrates supporting tool use. Cereb Cortex 20(3):507–516

Rilling JK, Glasser MF et al (2008) The evolution of the arcuate fasciculus revealed with comparative DTI. Nat Neurosci 11(4):426–428

Rizzolatti G, Arbib MA (1998) Language within our grasp. Trends Cogn Sci 21(5):188–194

Rizzolatti G, Luppino G, Matelli M (1998) The organization of the cortical motor system: new concepts. Electroencephalogr Clin Neurophysiol 106(4):283–296

Ruff CB, Trinkaus E, Holliday TW (1997) Body mass and encephalization in Pleistocene Homo. Nature 387(6629):173–176

Sagi Y, Tavor I et al (2012) Learning in the fast lane: new insights into neuroplasticity. Neuron 73(6):1195–1203

Schenker NM, Buxhoeveden DP et al (2008) A comparative quantitative analysis of cytoarchitecture and minicolumnar organization in Broca's area in humans and great apes. J Comp Neurol 510(1):117–128

Schlegel AA, Rudelson JJ et al (2012) White matter structure changes as adults learn a second language. J Cogn Neurosci 24(8):1664–1670

Schmahmann JD, Pandya DN et al (2007) Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. Brain 130(Pt 3):630–653

Scholz J, Klein MC et al (2009) Training induces changes in white-matter architecture. Nat Neurosci 12(11):1370–1371

Semaw S, Rogers MJ et al (2003) 2.6-Million-year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. J Hum Evol 45(2):169–177

Sisti HM, Geurts M et al (2012) Microstructural organization of corpus callosum projections to prefrontal cortex predicts bimanual motor learning. Learn Mem 19(8):351–357

Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17(3):143–155

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44(1):83–98

Smith SM, Jenkinson M et al (2004) Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23(Suppl 1):S208–S219

Smith SM, Jenkinson M et al (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31(4):1487–1505

Song SK, Sun SW, Ramsbottom MJ, Chang C, Russell J, Cross AH (2002) Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. Neuroimage 17:1429–1436

Song SK, Sun SW, Ju WK, Lin SJ, Cross AH, Neufeld AH (2003) Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. Neuroimage 20:1714–1722

Steele CJ, Scholz J et al (2012) Structural correlates of skilled performance on a motor sequence task. Front Hum Neurosci 6:289

Stout D (2002) Skill and cognition in stone tool production: an ethnographic case study from Irian Jaya. Curr Anthropol 45(3):693–722

Stout D (2010) The evolution of cognitive control. Top Cogn Sci 2(4):614–630

Stout D (2011) Stone toolmaking and the evolution of human culture and cognition. Philos Trans R Soc Lond B Biol Sci 366(1567):1050–1059

Stout D, Chaminade T (2007) The evolutionary neuroscience of tool making. Neuropsychologia 45:1091–1100

Stout D, Chaminade T (2012) Stone tools, language and the brain in human evolution. Philos Trans R Soc Lond B Biol Sci 367(1585):75–87

Stout D, Toth N et al (2008) Neural correlates of Early Stone Age tool-making: technology, language and cognition in human evolution. Philos Trans R Soc Lond B 363:1939–1949

Stout D, Passingham R et al (2011) Technology, expertise and social cognition in human evolution. Eur J Neurosci 33(7):1328–1338

Stout D et al (2014) Late Acheulean technology and cognition at Boxgrove, UK. J Archaeol Sci 41:576–590

Takeuchi H, Sekiguchi A et al (2010) Training of working memory impacts structural connectivity. J Neurosci 30(9):3297–3303

Taubert M, Draganski B et al (2010) Dynamic properties of human brain structure: learning-related changes in cortical areas and associated fiber connections. J Neurosci 30(35):11670–11677

Taubert M, Lohmann G et al (2011) Long-term effects of motor training on resting-state networks and underlying brain structure. Neuroimage 57(4):1492–1498

Taubert M, Villringer A et al (2012) Learning-related gray and white matter changes in humans: an update. Neuroscientist 18(4):320–325

Teffer K, Semendeferi K (2012) Human prefrontal cortex: evolution, development, and pathology. Prog Brain Res 195:191–218

Thiebaut de Schotten M, Dell'Acqua F et al (2012) Monkey to human comparative anatomy of the frontal lobe association tracts. Cortex 48(1):82–96

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

Tseng BY, Gundapuneedi T et al (2013) White matter integrity in physically fit older adults. Neuroimage 82C:510–516

Uomini NT, Meyer GF (2013) Shared brain lateralization patterns in language and acheulean stone tool production: a functional transcranial doppler ultrasound study. PLoS One 8(8):e72693

Vanduffel W, Fize D et al (2002) Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science 298(5592):413–415

Vigneau M, Beaucousin V et al (2011) What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing?: insights from a meta-analysis. NeuroImage 54(1):577–593

Wakana S, Caprihan A et al (2007) Reproducibility of quantitative tractography methods applied to cerebral white matter. Neuroimage 36(3):630–644

Wang Y, Adamson C et al (2012) Sex differences in white matter development during adolescence: a DTI study. Brain Res 1478:1–15

Weber BH, Depew DJ (eds) (2003). Evolution and learning: the Baldwin effect reconsidered. MIT Press, Cambridge

Wheeler-Kingshott CA, Cercignani M (2009) About "axial" and "radial" diffusivities. Magn Reson Med 61(5):1255–1260

Woolrich MW, Jbabdi S et al (2009) Bayesian analysis of neuroimaging data in FSL. Neuroimage 45(1 Suppl):S173–S186

Zatorre RJ, Fields RD, Johansen-Berg H (2012) Plasticity in gray and white: neuroimaging changes in brain structure during learning. Nat Neurosci 15(4):528–536

Zhang J, Jones M, DeBoy CA, Reich DS, Farrell JA, Hoffman PN, Griffin JW, Sheikh KA, Miller MI, Mori S et al (2009) Diffusion tensor magnetic resonance imaging of Wallerian degeneration in rat spinal cord after dorsal root axotomy. J Neurosci 29:3160–3171