Linking planning performance and gray matter density in mid-dorsolateral prefrontal cortex: Modulating effects of age and sex

Christoph P. Kaller, a,b,* Katharina Heinze, b,c Irina Mader, b,c Josef M. Unterrainer, d Benjamin Rahm, d Cornelius Weiller, a,b Lena Köstering, a,b

A R T I C L E   I N F O

Article history:
Accepted 12 August 2012
Available online 17 August 2012

Keywords:
Planning ability
Mid-dorsolateral prefrontal cortex
Voxel-based morphometry
Age effects
Sex differences

A B S T R A C T

Planning of behavior relies on the integrity of the mid-dorsolateral prefrontal cortex (mid-dlPFC). Yet, only indirect evidence exists on the association of protracted maturation of dlPFC and continuing gains in planning performance post adolescence. Here, gray matter density of mid-dlPFC in young, healthy adults (18–32 years) was regressed onto performance on the Tower of London planning task while accounting for moderating effects of age and sex on this interrelation. Multiple regression analysis revealed an association of planning performance and mid-dlPFC gray matter density that was especially strong in late adolescence and early twenties. As expected, for males better planning performance was linked to reduced gray matter density of mid-dlPFC, possibly due to maturational processes such as synaptic pruning. Most surprisingly, females showed an inverted, positive interrelation of planning performance and mid-dlPFC gray matter density, indicating that sexually dimorphic development of dlPFC continues during early adulthood. Age and sex are hence important moderators of the link between planning performance and gray matter density in mid-dlPFC. Consequently, the assessment of moderator effects in regression designs can significantly enhance understanding of brain-behavior relationships.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Successful completion of deliberate behavior beyond everyday routine relies on the ability to identify and select an appropriate sequence of actions before their actual execution. This ability to plan ahead future actions encompasses the mental conception and evaluation of several behavioral alternatives and their associated consequences (Goel, 2002; Ward and Morris, 2005). As one of the highest human cognitive abilities, it depends on the integrity of the prefrontal cortex (Owen, 2005). In particular, the functional contributions of the mid-dorsolateral part of the prefrontal cortex (mid-dlPFC) that refers to Brodmann areas 46 and 9/46 in middle frontal gyrus have been implicated to play a major role in planning and organization of behavior (e.g., Petrides, 2005; Unterrainer and Owen, 2006). Assessment of planning ability often employs disc-transfer paradigms such as the Tower of London task that was originally developed to examine planning impairments in patients with frontal lesions (Shallice, 1982). In this task, planning is required for an efficient transformation of a given start state into a predetermined goal state within the minimum number of moves (Berg and Byrd, 2002; Kaller et al., 2011a; for an illustration, see Fig. 1).

Previous developmental studies on the Tower of London task documented age-related improvements in planning ability from preschool age to adolescence (e.g., Asato et al., 2006; Huizinga et al., 2009; Kaller et al., 2011; Krikorian et al., 1994; Levin et al., 1991; Luciana and Nelson, 1998; Luciana and Nelson, 2002; Luciana et al., 2005). In a recent multi-center study employing an exceptionally large sample (n = 890, ranging from 10 to 30 years) and also more complex and demanding Tower of London problems with up to seven moves, Albert and Steinberg (2011) revealed that the developmental progress in planning ability extends beyond adolescence well into the mid-twenties (see also De Luca et al., 2003). In close parallel, mid-dlPFC is reported to manifest a likewise protracted course in its ontogenetic maturation that seemingly echoes its late evolvement during phylogenesis (Gogtay et al., 2004). Post-mortem investigations revealed not only a delayed synaptogenesis for middle frontal gyrus during early childhood, but also a considerable protraction in synapse elimination during adolescence (Huttenlocher and Dabholkar, 1997). In vivo neuroimaging studies further showed that, compared to adjacent structures in more anterior and posterior parts of prefrontal cortex, mid-dlPFC reaches its maximum cortical thickness at a later age during...
childhood and is even one of the latest peaking regions in the entire human brain (Gogtay et al., 2004; Shaw et al., 2008). Moreover, subsequent reduction in gray matter density particularly in dorsal parts of frontal cortex proceeds well into the post-adolescent mid-twenties (Sowell et al., 1999; Sowell et al., 2001), possibly reflecting anatomical refinements of neuronal processing via synaptic pruning and the elimination of excess connections (Giedd, 2004; Gogtay et al., 2004; but see also Paus, 2005). In this regard, a recent study on post-mortem brain tissue confirmed that elimination of synaptic spines in dlPFC continues throughout the third decade of life (Petanjek et al., 2011).

Based on this apparent coincidence between (i) the prolonged cognitive development of planning ability and (ii) the protracted brain maturation in (mid-)dorsolateral prefrontal cortex that both extend well into adulthood, Albert and Steinberg (2011) suggested a direct link between these two processes. However, although intriguing and in line with extant views on the development of prefrontal function (e.g., Blakemore and Choudhury, 2006; Casey et al., 2005; Diamond, 2002; Paus, 2005), the specific link for planning ability has not been established in anatomical data so far.

In the present study, we therefore investigated the relationship between gray matter density in mid-dlPFC and planning ability in a larger sample of late adolescents and young adults (18–32 years). Region-of-interest (ROI) analyses were based on functionally defined locations of left and right mid-dlPFC derived from a previous experiment employing the Tower of London task in combination with functional magnetic resonance imaging (Kaller et al., 2011b). Given that the parallel of developmental changes in morphology and cognitive ability further suggests a moderating effect of age on the presumed brain-behavior relationship (also see below), multiple regression analysis with interaction effects was applied (cf. Cohen et al., 2003; Jaccard and Turrisi, 2003). Moreover, putative sex-related differences in Tower of London performance remain a matter of contention (e.g., Boghi et al., 2006; De Luca et al., 2003; Unterrainer et al., 2005), while sexual dimorphisms in brain development and resultant morphological organization are commonly known (e.g., Allen et al., 2003; Giedd et al., 1999; Goldstein et al., 2001; Gur et al., 2002; Lenroot et al., 2007). Hence, besides age, sex was added as another moderator to account for potential sex-related variability in the relationship between planning performance and gray matter density in mid-dlPFC.

Methods

Subjects

Present analyses were based on two previously acquired data sets (incl. pilots) that both comprised the administration of an identical problem set of the Tower of London planning task as well as the acquisition of anatomical data using an identical imaging protocol on the same magnetic resonance imaging (MRI) scanner. In the first study (Kaller et al., in press), general planning ability was initially assessed with a standard four- to six-move Tower of London problem set as means for controlled assignment of subjects to different experimental groups, that were then tested using a different variant of the task during transcranial magnetic stimulation (TMS). Only data of the initial testing with the original Tower of London were included in the present analyses, along with anatomical MRI scans acquired for neuronavigated application of stimulation over mid-dlPFC. The sample included here (Sample 1, n=59, 26 female; age M = 24.24 years, SD = 2.51; all right-handed) comprised the resulting data sets after exclusion of two subjects that had been also excluded in the study of Kaller et al. (in press) either due to an incidental finding in the anatomical MRI scan or severely poor performance in the initial assessment of global planning ability. Contrary to the data set of Kaller et al. (in press), the current sample included a TMS pilot subject who was pre-tested but then stimulated using a different TMS protocol. In addition, the present sample set included six data sets that had been discarded in the TMS study due to events during the TMS session (technical malfunctions, safety reasons, non-compliant behavior, outlier), as these reasons were unrelated to the initial assessments of global planning ability considered here.

In the second study (Kaller et al., 2012), psychometric properties of an extended version of the same standard Tower of London problem set were assessed. Due to participation in other imaging studies, anatomical MRI scans were available for a subgroup of this study’s subjects (Sample 2, n=45, 19 female; age M = 23.56 years, SD = 2.76; all right-handed). Given that (i) the subjects included from these two sources showed a virtually identical planning accuracy in terms of proportions of problems perfectly solved in the minimum number of moves (Sample 1, M = 72.67%; SD = 9.83; Sample 2, M = 71.34%; SD = 11.67; t(102) = .607, p = .545) and that (ii) they also did not differ with respect to age (t(102) = 1.324, p = .188) or sex (χ² = .035, df = 1, p = .851), the two data sets were collapsed in the subsequent analyses (for additional assessments of the data sets’ comparability, see also Tower of London task section).

In total, data of 104 subjects were included in the present analyses (45 female). Subjects were aged between 18 and 32 years (M = 23.95 years, SD = 2.63). All subjects were right-handed and had normal or corrected-to-normal visual acuity. None of them was under medical treatment or reported a history of psychiatric or neurological illness. In both primary studies on the Tower of London task, written informed consent was obtained prior to participation (cf. Kaller et al., in press, 2012). In addition, acquisition of anatomical images was approved by local ethics authorities. Subjects received monetary compensation for their participation (approximately 10 €/h).

Tower of London task

The Tower of London task (Shallice, 1982) is a frequently used neuropsychological test instrument for assessing planning ability in various clinical and healthy populations (Kaller et al., 2011a). In its original version, three balls of different colors are placed on three different rods of different lengths (Berg and Byrd, 2002). Subjects are presented with a start state and are instructed to transform it into a given goal state. In order to solve the problem in the least possible number of moves, subjects are thus requested to plan ahead a solution before manually executing the moves. Three rules have to be followed: (i) Only one ball can be moved at a time, (ii) balls must not be placed outside the tower, and (iii) if more than one ball is stacked on a rod, only the topmost ball can be moved.

The Tower of London problem set applied in the two primary studies consisted of an optimized problem selection recently suggested by Kaller et al. (2011a). In its extended version this problem set comprises eight four-, five-, six-, and seven-move problems each that instantiate a linear increase of problem difficulty (see Kaller et al., 2012, for detailed psychometric evaluations). For the present analysis, however, only four-, five- and six-move problems were considered because subjects in Sample 1 were not assessed with seven-move problems (cf. Kaller et al., in press).

In both samples, subjects were tested individually in quiet laboratory rooms and administered a computerized version of the original Tower of
London. The computer program did not allow rule-incongruent moves. Start state and goal state were presented in the lower and upper part of the screen, respectively. The minimum number of moves for the current trial was indicated on the left side of the start state. Problem presentations were carried out on 17" monitors; the task's physical appearance was identical across samples, and the presentation of a single trial was limited to 1 min (cf. Shallice, 1982). In both studies, written and oral instructions placed strict emphasis on planning ahead before starting any move execution.

The two primary studies differed only with respect to the response mode: Subjects in Sample 1 moved individual balls from/to different pegs by clicking on the corresponding buttons of a three-button computer mouse (cf. Keller et al., in press) whereas subjects in Sample 2 executed ball movements on a touch-sensitive screen (cf. Keller et al., 2012). In order to exclude the possibility that the different response modes affected subjects' planning by any means, initial thinking and movement execution times were entered as dependent variables into separate repeated-measures ANOVAs with minimum moves as within-subjects factor and sample as between-subjects factor. As expected, movement execution times differed significantly between samples \((F_{1,102} = 8.85, p = .001\); interaction: \(F_{1,102} = 4.69, p = .030\). But most importantly, initial thinking times showed neither a main effect of sample \((F_{1,102} = .21, p = .645\) nor an interaction \((F_{1,102} = .10, p = .909)\). Note that in addition to the above reported t-test on overall planning accuracy, comparing the samples' accuracies as a function of minimum moves did also not reveal any differences in terms of an interaction \((F_{1,102} = 1.17, p = .330\); main effect of sample \(F_{1,102} = 36.8, p = .545\). Thus, with respect to the assessment of core planning processes, the two primary studies can be regarded as comparable, again demonstrating that the combination of the two data sets was justified (see also Subjects section).

As the main variable of interest here, planning accuracy was quantified as the proportion of problems correctly solved (i.e., within the minimum number of moves) without exceeding the time limit of 1 min (see above). Analyses of sample descriptive and behavioral data were conducted using SPSS Statistics 20 (IBM Corp., Armonk, NY, U.S.A.).

Anatomical brain imaging

All included subjects underwent T1-weighted high-resolution anatomical brain imaging on the same 3 T TIM TRIO whole-body MRI scanner (SIEMENS, Erlangen, Germany) using magnetization-prepared rapid gradient echo (MPRAGE) imaging with the following scan acquisition parameters: repetition time, 2200 ms; echo time, 2.15 ms; inversion time, 1100 ms; flip angle, 12°; 160 sagittal slices; matrix size, 256×256; field of view, 256 mm, resulting in 1.0 mm³ cubic voxels. The time interval between accomplishment of the Tower of London task and acquisition of anatomical images was 0.35 years on average \((SD = .54, maximum 1.75 years)\).

Preprocessing of structural images and voxel-based morphometry (VBM)

Structural images were processed in MATLAB 7.9.0 (R2009b; The Mathworks, Inc., Natick, MA, U.S.A.) using the SPM8 (release r4667; http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) and VBM8 (release r435; http://dbm.neuro.uni-jena.de/) software packages. At first, anatomical images were manually reoriented placing the coordinate system’s origin into the anterior commissure and the posterior commissure into the intersecting xy-plane. Segmentation of individual brains into compartments of gray and white matter and cerebrospinal fluid was based on VBM8 using the default estimation options (i.e., very light bias regularization, 60 mm cut-off for estimating the Gaussian smoothness of bias in image intensity; ICBM [International Consortium for Brain Mapping] European template for initial affine transformation) and tissue probability maps (modified version of ICBM tissue probabilistic atlases). Spatial normalization into the Montreal Neurological Institute (MNI) standard space was done by the high-dimensional DARTEL (diffeomorphic anatomical registration through exponentiated lie algebra; see Ashburner, 2007) approach implemented in VBM8. To this end, default options such as multi-threaded SANLM (spatial adaptive non local means) denoising, light clean-up of partitions, and weighting of Hidden Markov–Random Fields by a factor of .15 were used. As suggested in VBM8, normalized probability maps of individual subjects’ gray matter density were written with non-linear but not affine modulation which effectively results in a multiplicative correction of the data for different brain sizes across subjects. Modulated normalized images were spatially smoothed with a kernel of 12 mm full width at half maximum (FWHM). Finally, smoothed images were inspected for poor image quality using the sample homogeneity tool implemented in VBM8 that computes the covariance of all images across the sample. Images of two male subjects that deviated more than two standard deviations from the mean covariance were carefully checked for potential artifacts, particularly in dorsolateral prefrontal cortex. As no abnormalities could be identified, all subjects were included in the subsequent analyses. Total intracranial volume was computed as the sum of the segmented volumes for gray matter, white matter, and cerebrospinal fluid.

Region-of-interest specification and data extraction

Effect sizes for the presumed relationship between gray matter density and planning accuracy were expected to be small. In consequence, regions-of-interest (ROI) were selected in mid-dLPFC based on a previous fMRI activation study that revealed dissociable contributions of left and right mid-dLPFC in planning (Kaller et al., 2011b; see also Keller et al., in press). Whereas activation contrasts reported in this study focused on differences in task-demand dependent lateralization in mid-dLPFC, here we re-assessed fMRI activation associated with planning in general, i.e. irrespective of experimental manipulation of planning demands. The whole-brain analysis (corrected for family-wise error, FWE) revealed a cluster in left and right mid-dLPFC (peak coordinates in MNI stereotactic space: \([-36 \, 38 \, 28]\) and \([36 \, 40 \, 34]\), respectively). For subsequent regression analyses on gray matter density in mid-dLPFC, two spherical ROIs with a radius of 12 mm were specified based on these coordinates (see Fig. 4A). Individual estimations of gray matter density were extracted from the smoothed, modulated, and (non-linear but not affine) normalized probability maps of gray matter preprocessed with VBM8 (see Preprocessing of structural images and voxel-based morphometry (VBM) section) and aggregated using the first eigenvariate from the principal component analysis approach implemented in SPM8. For extraction, a masking procedure assured that only those voxels within the ROIs were included that had a minimum probability of 20% for gray matter in all subjects.

Multiple regression analyses with interaction effects

Although it is common usage to assess non-additive effects – that is, interactions – in factorial designs, this approach is only rarely applied in the context of multiple regression analyses. In multiple regression, non-additive effects are accounted for by extending the additive (or “main effects”) model with product terms reflecting the non-additive components (for introductions, see Cohen et al., 2003; Jaccard and Turrisi, 2003). In the context of the present analyses, it was proposed that gray matter density in mid-dLPFC would be related to planning accuracy...
with lower density being associated with better performance. Yet, given that mid-dlPFC is one of the latest maturing structures of the brain with protracted brain development (e.g. synaptic pruning) proceeding well into post-adolescence (Cogtay et al., 2004; Huttenlocher and Dabholkar, 1997; Shaw et al., 2008; see also Badre and D’Esposito, 2009), it was further proposed that the relationship between gray matter density and planning accuracy (as the focal variable of interest here) would be moderated by age. Specifically, it was expected that the extent of the negative correlation between gray matter density and planning accuracy varied systematically with age and was more pronounced in late adolescence/early twenties than in late twenties. This assumption was based on the fact that inter-individual variation in gray matter density and planning accuracy in early twenties should be due to both developmental differences in brain maturation as well as general differences in intellectual capabilities whereas it should most likely reflect only the latter source of variance in late twenties. In order to account for potential morphological differences between male and female subjects (e.g., Allen et al., 2003; Goldstein et al., 2001; Gur et al., 2002; Lenroot et al., 2007), present analyses were further complemented by sex as factor. Thus, the resulting regression model comprised gray matter density as dependent variable Y, the focal variable planning accuracy (TOL), the first- and second-order moderator variables age and sex, and all two- and three-way interactions.

It is important to note that in the framework of multiple regression with interaction effects, resulting test statistics (e.g. regression coefficients, respective T statistics) are conditional and their specific values are only valid for a given centering of the other variables (Cohen et al., 2003; Jaccard and Turrisi, 2003). In accordance, different models with a systematically shifted centering of the moderator variables are computed in order to conceive the overall pattern, whereas the focal variable (here TOL) is kept mean-centered. In the present analyses, the impact of the continuous first-order moderator variable age was observed in three different centerings based on mean and standard deviation of subjects’ age (i.e. planning accuracy, TOL) and the two moderator variables (age, sex) as covariates in the above ANOVA on planning accuracy did also not reveal any interactions with minimum moves (highest F = .857, lowest p = .490). Thus, present focal and moderator variables were devoid of any collinearity. Analyses of total intracranial volume did also not yield any correlations with planning accuracy (r = .103, p = .299) or age (r = .120, p = .223), but of course revealed significantly larger volumes in males compared to females (t(102) = 8.76, p < .001). Bivariate distributions of planning accuracy, age, sex, and total intracranial volume are illustrated in Fig. 3.

**Regression analyses on gray matter density in mid-dlPFC**

Multiple regression analyses on gray matter density were computed separately for left and right mid-dlPFC ROIs (cf. Fig. 4A) and included three regressors for the simple effects\(^4\) of the focal variable of interest (i.e. planning accuracy, TOL) and the two moderator variables (age, sex) as well as four additional regressors for all possible two- and three-way interactions (TOL+age, TOL+sex, age+sex, TOL+age+sex) and a constant term for the intercept (see Methods in Multiple regression analyses with interaction effects section). All regressors were forced to be entered into the regression model.

Above all, results revealed a significant three-way interaction for TOL+age+sex in left mid-dlPFC (t(102) = -2.32, p = .023) that was mirrored by a trend for the same effect in right mid-dlPFC (t(102) = 1.77, p = .080). Thus, the extent of the relation between planning accuracy and gray matter density in left and right mid-dlPFC was moderated by age and sex. In other words, the slopes for regressing gray matter density on planning accuracy varied depending on both subjects’ age and sex. This becomes evident also in the overview on the single-regressor test statistics provided in Tables 1 and 2 that were computed for different centerings of the moderator variables age and sex (see Methods in Multiple regression analyses with interaction effects section). The resulting regression lines for predicting gray matter density from planning accuracy as a function of age and sex are illustrated in Figs. 5A and B.

As is visible from comparing sex-specific regression slopes in Figs. 5A and B, the moderating effect of age on the relationship between TOL and mid-dlPFC gray matter density differed between

---

\(^{2}\) The IFX toolbox is a custom collection of MATLAB scripts that facilitate multiple regression analyses with interaction effects in SPMB. In particular, the scripts automatize the computation of different models with shifted centerings of moderator variables as well as the subsequent integration and visualization of effects. Readers interested in using the toolbox are asked to contact the corresponding author for a copy.

\(^{3}\) Model 3 of the PROCESS macro was used specifying left and right mid-dlPFC volume as outcome variable Y, respectively, TOL as focal predictor X, AGE as first-order moderating variable M, and SEX as second-order moderator W (Hayes, 2012; see also www.afhayes.com).

\(^{4}\) In contrast to classical main effects in factorial designs that represent the effect of an independent variable averaged across all levels of other independent variables, simple effects assess the bivariate relationship between a predictor and the outcome variable under the condition that the regression weights of the other regressors are set to zero (Jaccard and Turrisi, 2003). That is, simple effects represent the effect of a predictor at specific values of the other predictors.
male and female subjects. For males, planning accuracy was mainly negatively associated with gray matter density in both left and right mid-dlPFC. In addition, this association was moderated by age in that the negative interrelation was strongest for young males (evaluated with the age regressor centered on M-SD, i.e. 21.32 years), but decreased with advancing age (evaluated with the age regressor centered on M and M + SD, respectively, i.e. 23.95 and 26.58 years). Thus, the negative relationship between planning accuracy and mid-dlPFC gray matter density was especially pronounced for males in late adolescence/young adult age (see black lines in left-hand columns of panels A and B of Fig. 5). For females, the direction of the relationship between planning accuracy and left and right mid-dlPFC gray matter density was inverted compared to males. That is, planning accuracy was positively associated with gray matter density of mid-dlPFC. This association was likewise moderated by age in that the positive interrelation was strongest for young females (evaluated with the age regressor centered on M-SD, i.e. 21.32 years; see black lines in right-hand columns of panels A and B of Fig. 5) and decreased with advancing age (evaluated with the age regressor centered on M and M + SD, respectively, i.e. 23.95 and 26.58 years). This age- and sex-specific pattern also becomes apparent by the fact that the two-way interactions TOL × sex indicating a different relation between planning accuracy and gray matter density for males versus females were only significant at the centering for younger ages (left mid-dlPFC, t_{102} = 2.43, p = .017; right mid-dlPFC, t_{102} = 2.36, p = .020).

For further exploring the regional specificity of this moderator effect of age and sex on the relation between planning accuracy and gray matter density within the dorsolateral part of prefrontal cortex, voxel-wise statistics of the underlying three-way interaction were rendered on the overall sample’s average brain. That is, equal to common statistical parametric maps, a spatial map was created indicating which voxels emerged as significant for the three-way interaction effect of TOL × age × sex. For this purpose, a mask for the dorsolateral prefrontal cortex was generated using the WFU PickAtlas version 3.0.3 (Maldjian et al., 2003; http://fmri.wfubmc.edu/software/PickAtlas/) by combining bilateral segments for middle and superior frontal gyrus derived from the IBASPM116 segregation. Voxel-wise p-values of the three-way interaction effect within this dlpFC mask were then rendered on the average brain of the overall sample. As illustrated in Fig. 4B, the resulting maps revealed that the three-way interaction of TOL × age × sex in predicting voxel-based morphometric differences in gray matter density was relatively focused to mid-dlPFC in both hemispheres, thus strikingly overlapping with previous planning-related fMRI activations in left and right mid-dlPFC (cf. Kaller et al., 2011b; see also Fig. 4, panel A). That is, the circumscribed region for which differences in gray matter density could be significantly predicted here by planning accuracy, under moderating effects of age and sex, closely overlaps with the region that was revealed to be significantly activated during planning in a previous fMRI study (Kaller et al., 2011b), thus indicating that planning accuracy is consistently linked both to the structure and function specifically of the middle part of dlPFC.

Supplementary analyses

In line with previous findings (e.g. Gur et al., 2002; Lenroot et al., 2007; Luders et al., 2009; Sowell et al., 2007), male and female subjects of the present study also differed substantially with respect to global brain volume (see Fig. 3). In order to preclude that the present three-way interaction (including sex) was driven by the chosen multiplicative correction for differences in brain size (using non-linear but not affine transformation parameters for modulation; see Methods in Preprocessing of structural images and voxel-based morphometry (VBM) section), analyses were repeated applying an additive approach to account for differences in brain volume (by including subjects' total intracranial volume into the statistical model after application of non-linear and affine transformation parameters for modulation). Results replicated the present findings by yielding a significant three-way interaction in left mid-dlPFC (t_{102} = −2.41, p = .018) and a trend in right mid-dlPFC (t_{102} = −1.79, p = .076). Moreover, analyses were repeated also without any modulation, again showing a significant three-way interaction in left mid-dlPFC (t_{102} = −2.01, p = .047), but not in right mid-dlPFC (t_{102} = −1.16, p = .247). Detailed results of all these analyses are reported in the Supplementary Materials and further extended in the Discussion.

Discussion

The present analyses revealed two main results: First and foremost, planning accuracy and gray matter density in mid-dlPFC were interrelated, with the strength of this association being moderated by age. That is, gray matter density in mid-dlPFC differentiated between high- and low-performing planners in late adolescence and the early twenties, whereas this interrelation declined until the late twenties. Second, the specific pattern of this moderator effect of age was entirely different for both sexes. In line with expectations, young males demonstrated a negative association between planning accuracy and mid-dlPFC gray matter, possibly indicating an association between protracted maturational processes such as synaptic pruning and enhanced efficiency of cognitive functioning. For females, a highly unusual pattern was observed suggesting a positive relationship between planning accuracy and gray matter density of dlPFC — again most pronounced in the early twenties. Thus, strikingly inverted sex-specific patterns of the associations between gray matter density of mid-dlPFC and planning accuracy were found for males and females particularly in late adolescence and early post-adolescence.

The mid-dlPFC is unequivocally regarded as critical for planning and deliberate organization of behavior (e.g., Petrides, 2005; Untrrainer and Owen, 2006), as has been frequently demonstrated in functional...
imaging studies employing the Tower of London task (e.g., Kaller et al., 2011b; Lazoner et al., 2000; Schall et al., 2003; Unterrainer et al., 2004; van den Heuvel et al., 2003). Thus far, only indirect evidence existed marking early post-adolescence as a prolonged critical stage for development of planning ability and its underlying neuronal structures (Albert and Steinberg, 2011). Current results, however, present first evidence for a direct link between differences in mid-dlPFC morphology and planning ability in late and post-adolescence. In males, the direction of this association was negative, suggesting that relatively more gray matter in mid-dlPFC was informative of less efficient planning. This can be well integrated into extant views on brain maturation which posit that processes of synaptic pruning and elimination of excess neuronal connections in adolescence enhance neural processing efficiency and that these processes are protracted into post-adolescence (mid-dlPFC), rather than the dlPFC as a whole, is functionally (cf. Kaller et al., 2011b) as well as structurally linked to planning accuracy. Abbreviations: L, left; R, right; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; SFG, superior frontal gyrus.

Surprisingly, however, females significantly differed from males by demonstrating an inverted, positive association between planning accuracy and mid-dlPFC gray matter density. This is seemingly at odds with general brain development which evidences a decrease in gray matter after peak densities are reached in early adolescence exhibiting a pattern suggestive of enhanced planning efficiency due to reductions in excessive gray matter density of mid-dlPFC.

It has to be noted that for the present sample age did not correlate with Tower of London performance. That is, the proposed higher efficiency of neuronal processing was not reflected by better behavioral planning accuracy in the older post-adolescent subjects. Yet, higher efficiency does not necessarily become apparent in behavioral data, as it may instead lead to recruitment of smaller neuronal ensembles underlying functional organization or the preference of problem solving strategies, or a different recruitment that is secondary to morphological functioning. Thus, the question is raised as to which factors underlie these counterintuitive findings for females. From a functional perspective, Boghi et al. (2006) found stronger activation in bilateral mid-dlPFC for females compared to males in complex Tower of London six-move problems indicating either sexual differences in the underlying functional organization or the preference of problem solving strategies, or a different recruitment that is secondary to morphological differences (but see also Unterrainer et al., 2005). As for the latter, the dlPFC is an area known to exhibit sexual differences in gray matter density, typically with larger volumes in females (Goldstein et al., 2001; Schlaepfer et al., 1995; Pietzer et al., 2010; but see Gur et al., 2002). Furthermore, the age-related trajectory of frontal cortical maturation is known to be sexually dimorphic with males exhibiting a delayed peak of gray matter density (Giedd et al., 1999; Lenroot et al., 2007) and a

---

Fig. 4. (A) Spatial localization of the applied spherical regions-of-interest (ROIs) with 12 mm radius in left (−36 38 28) and right (36 40 34) mid-dorsolateral prefrontal cortex (mid-dlPFC). Anatomical demarcation of extracted mid-dlPFC gray matter was directly based on location of task-related activity in a previous fMRI study (Kaller et al., 2011b). (B) For illustrative purposes, voxel-wise statistics of the three-way interaction (i.e. the moderating effect of age and sex on the relation between gray matter density and planning accuracy) were rendered within dorsolateral prefrontal cortex on the overall sample’s average brain. The overlay shows that the voxels where the present three-way interaction reached significance highly overlap with ROI coordinates of previous mid-dlPFC activations displayed in panel A. This indicates that specifically the middle part of the dorsolateral prefrontal cortex (mid-dlPFC), rather than the dlPFC as a whole, is functionally (cf. Kaller et al., 2011b) as well as structurally linked to planning accuracy. Abbreviations: L, left; R, right; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; SFG, superior frontal gyrus.

---

5 Illustrations in Figure 5 indicate that the relation between planning accuracy and gray matter density seems not only to decline across age but to slightly reverse in direction (see regression lines for centerings at older ages, M±SD). However, given the present restrictions to strictly linear relations and previous reports of higher-order non-linear patterns of longitudinal developments in this age range (e.g., Giedd et al., 1999; Shaw et al., 2008), this seeming inversion of slopes should be interpreted with great caution.
Steeper post-adolescent decline in volume of middle frontal gyrus (Taki et al., 2011; also see Sowell et al., 2001) and in cortical thickness specifically of dlPFC (Raznahan et al., 2010) compared to females. That is, sex differences in age-related effects on the relationship of planning accuracy and mid-dlPFC gray matter density could in part have been driven by cross-sectionally sampling different developmental stages of protracted maturation of the dlPFC in young females versus males of comparable chronological age.

The aforementioned explanation, however, would still not account for the inversion of the association of mid-dlPFC gray matter density and planning accuracy in females compared to males. In this regard, one possible factor differentially affecting brain morphology in females might be their increased sensitivity to the effects of estrogen on brain structure. Gonadal hormones are generally known to affect brain organization (e.g., Raznahan et al., 2010; Witte et al., 2010; for a review, see Peper et al., 2011) and the middle frontal gyrus is a region with high density of sex hormone receptors during developmentally critical stages (Goldstein et al., 2001). Accordingly, increased estrogen levels were found to be significantly related to increased volume of the middle frontal gyrus in pubertal girls, but not boys (Peper et al., 2009). Furthermore, women using hormonal contraceptives, which typically enhance estrogen and progesterone levels, have recently been found to have increased volume of the bilateral dlPFC and adjacent lateral prefrontal areas compared to naturally cycling women (Pletzer et al., 2010). In a similar vein, evidence from primate research showed that estrogen exerted stimulatory effects over density of catecholamine fibers in dlPFC of female adult monkeys (Kritzer and Kohama, 1998), thereby possibly modulating cognitive functions dependent on dopaminergic transmission in prefrontal cortex. Likewise, it has recently been established that estrogen bears functional relevance for cognitive performance and its underlying prefrontal dopaminergic activity in humans (Jacobs and D'Esposito, 2011). When performing the N-Back working memory task, naturally cycling young women with low estradiol levels demonstrated increased, less efficient task-related mid-dlPFC activity, which was driven by reduced prefrontal dopamine availability, compared to women with high estradiol levels (Jacobs and D'Esposito, 2011). Thus, cumulating evidence points towards a differential influence of female gonadal hormones on prefrontal morphology and function. Although purely speculative, it is possible that in those females of the current study, that were still in a developmentally sensitive stage of dlPFC maturation, variations in estrogen levels were related to increased mid-dlPFC gray matter density and to enhanced efficiency of prefrontal activity on which the Tower of London is known to critically rely, thereby producing the counterintuitive positive association of planning accuracy and gray matter density in mid-dlPFC. Again, as neither estradiol levels nor use of hormonal contraceptives were recorded, it has to be stressed that this poses a possible, but purely speculative explanation. In any case, the sexually dimorphic age-effects on the relation between dlPFC morphology and planning performance represent an intriguing finding with potential ramifications for existing views on sex differences in brain-behavior relationships if replicated in future studies. In this respect, an interesting finding

---

**Table 1**

<table>
<thead>
<tr>
<th>Centering of moderator variables</th>
<th>Unstandardized regression coefficients</th>
<th>r scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M – SD</td>
<td>M + SD</td>
</tr>
<tr>
<td></td>
<td>(23.95–2.32 years)</td>
<td>(23.95+2.32 years)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Regressors</td>
<td>Intercept</td>
<td>55.0513</td>
</tr>
<tr>
<td></td>
<td>TOL</td>
<td>−0.2806</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>−0.3433</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>−1.3278</td>
</tr>
<tr>
<td></td>
<td>TOL + age</td>
<td>0.0958</td>
</tr>
<tr>
<td></td>
<td>TOL + sex</td>
<td>1.2704</td>
</tr>
<tr>
<td></td>
<td>Age + sex</td>
<td>0.4056</td>
</tr>
<tr>
<td></td>
<td>TOL + age + sex</td>
<td>−0.3481</td>
</tr>
</tbody>
</table>

N.B. Significances r > 1.6608, p < 0.10; r > 1.9450, p < 0.05; r > 2.6280, p < 0.01, highlighted in bold font.

---

**Table 2**

<table>
<thead>
<tr>
<th>Centering of moderator variables</th>
<th>Unstandardized regression coefficients</th>
<th>r scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M – SD</td>
<td>M + SD</td>
</tr>
<tr>
<td></td>
<td>(23.95–2.32 years)</td>
<td>(23.95+2.32 years)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Regressors</td>
<td>Intercept</td>
<td>57.3318</td>
</tr>
<tr>
<td></td>
<td>TOL</td>
<td>−0.8675</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>−0.5044</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>−2.2884</td>
</tr>
<tr>
<td></td>
<td>TOL + age</td>
<td>0.1714</td>
</tr>
<tr>
<td></td>
<td>TOL + sex</td>
<td>1.6541</td>
</tr>
<tr>
<td></td>
<td>Age + sex</td>
<td>0.8627</td>
</tr>
<tr>
<td></td>
<td>TOL + age + sex</td>
<td>−0.3564</td>
</tr>
</tbody>
</table>

N.B. Significances r > 1.6608, p < 0.10; r > 1.9450, p < 0.05; r > 2.6280, p < 0.01, highlighted in bold font.
was recently reported by Zuo et al. (2010) who observed opposite developmental patterns for males and females in the functional homotypy of corresponding regions in left and right dlPFC using resting-state fMRI. A significant interaction revealed that functional connectivity of homologue parts in dlPFC increased across age in males but decreased in females, indicating sexually diverging trajectories in organizational principles such as functional segregation. In a similar regard, it has to be noted that – based on visual inspection – present sex-specific results did not seem to emerge equally strong for left and right mid-dlPFC (cf. Fig. 5, Table 1 and 2). However, from the statistical design applied here, no inferences can be drawn on the lateralization of structural differences between females and males so that this observation remains an open question for future studies.

Several methodological issues warrant caution when considering potential explanations of the present three-way interaction. For instance, males and females were not matched on the distribution of genetic polymorphisms that are known to modulate prefrontal functions and – in addition to hormonal differences – may have also driven the present differential findings (e.g. Wei et al., 2012). Furthermore, as often reported, males and females differ substantially with respect to brain sizes and volumes (cf. Fig. 3; see also Gur et al., 2002; Lenroot et al., 2007; Luders et al., 2005; Sowell et al., 2007). Although male brains are on average larger in terms of absolute size (e.g., Sowell et al., 2007), males are not attributable to a smaller brain size per se, but constitute a sex-dependent redistribution of tissue volume (Luders et al., 2009). Nonetheless, considering this possibility that present results were biased by normalizing the mostly smaller female brains and the mostly larger male brains into a standard space of average brain size.

Despite these methodological imponderables, applying multiple regression analysis with interaction effects to functional and structural magnetic resonance imaging data presents a fruitful and promising approach, because it can considerably extend knowledge on the dynamics...
of brain-behavior relationships. Conventional regression designs acknowledge the multi-factorial nature of brain-behavior relationships by commonly modeling multiple regressors. However, it is not plausible why regression designs should omit assessing how these factors interact with each other in predicting brain structure or function, whereas factorial designs routinely assess these non-additive effects. In particular, developmental and aging studies, which typically investigate how age and sex influence changes in brain morphology and related cognitive function, could benefit from knowledge on the interplay of these important determinants of brain-behavior relationships. Thus, with the current study we also hope to promote and facilitate the use of interactional regression designs (cf. Cohen et al., 2003; Jaccard and Turrisi, 2003), a viable avenue of research which has hitherto been seldom followed.

Conclusion

In sum, the statistical approach of modeling interaction effects in multiple regression applied here revealed unique associations between gray matter density of mid-dLPFC and planning ability in young healthy adults. Age is a significant moderator of this association in that planning accuracy of adults in late adolescence and early post-adolescence was significantly linked to gray matter density in mid-dLPFC, whereas this interaction declined for adults in their mid- to late twenties. That is, only in developmentally critical stages of ongoing maturation of executive functions and dorsolateral prefrontal brain regions, a direct link between better Tower of London performance and reduced dLPFC gray matter density exists. Whereas males exhibited an expected negative pattern for this interaction, suggesting that maturational processes such as synaptic pruning underlie efficient planning ability, for females better planning performance was related to increased gray matter density of dLPFC. To conclude, age and sex were revealed to be important moderators of the association between planning ability and the mid-dorsolateral prefrontal cortex. Explicitly addressing non-additive effects of age and sex can thus considerably extend understanding of the relationship between higher-order cognitive functions such as planning ability and brain morphology.

Acknowledgments

This work was supported by grants of the German Federal Ministry of Education and Research (BMBF; grant number 01GW0710) and the European Union FP 7 research program (PLASTICISE: Collaborative Project 223524).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.08.052.

References


