Developmental Cognitive Neuroscience 19 (2016) 211–222

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: http://www.elsevier.com/locate/dcn

Longitudinal development of frontoparietal activity during feedback learning: Contributions of age, performance, working memory and cortical thickness

Sabine Peters[a,b,*], Anna C.K. Van Duijvenvoorde[a,b], P. Cédric M.P. Koolschijn[c], Eveline A. Crone[a,b]

[a] Department of Developmental Psychology, Leiden University, The Netherlands
[b] Leiden Institute for Brain and Cognition, The Netherlands
[c] Dutch Autism & ADHD Research Center, Department of Brain and Cognition, University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 26 November 2015
Received in revised form 16 March 2016
Accepted 10 April 2016
Available online 13 April 2016

Keywords:
Feedback learning
Adolescence
Frontal cortex
Parietal cortex
Cognitive control
Development,

ABSTRACT

Feedback learning is a crucial skill for cognitive flexibility that continues to develop into adolescence, and is linked to neural activity within a frontoparietal network. Although it is well conceptualized that activity in the frontoparietal network changes during development, there is surprisingly little consensus about the direction of change. Using a longitudinal design (N = 208, 8–27 years, two measurements in two years), we investigated developmental trajectories in frontoparietal activity during feedback learning. Our first aim was to test for linear and nonlinear developmental trajectories in dorsolateral prefrontal cortex (DLPFC), superior parietal cortex (SPC), supplementary motor area (SMA) and anterior cingulate cortex (ACC). Second, we tested which factors (task performance, working memory, cortical thickness) explained additional variance in time-related changes in activity besides age. Developmental patterns for activity in DLPFC and SPC were best characterized by a quadratic age function leveling off/peaking in late adolescence. There was a linear increase in SMA and a linear decrease with age in ACC activity. In addition to age, task performance explained variance in DLPFC and SPC activity, whereas cortical thickness explained variance in SMA activity. Together, these findings provide a novel perspective of linear and nonlinear developmental changes in the frontoparietal network during feedback learning.

1. Introduction

The ability to learn from performance feedback is crucial to flexibly adapt to a changing environment. Behavioral performance during feedback learning shows a protracted development which continues into adolescence (Huizinga et al., 2006). Several studies have investigated the neural underpinnings of feedback processing. Studies in adults have shown that learning from feedback is associated with activity in a frontoparietal network, including dorsolateral prefrontal cortex (DLPFC), supplementary motor area (SMA), anterior cingulate cortex (ACC) and superior parietal cortex (SPC) (Carter and van Veen, 2007; Mars et al., 2005; Zanolie et al., 2008). Intriguingly, developmental neuroimaging studies have reported age-related activity changes in this network during feedback processing, suggesting an important link between feedback learning and neural maturation of the frontoparietal network (Crone et al., 2008; Peters et al., 2014a; Van Duijvenvoorde et al., 2008; Velanova et al., 2008). Despite these findings, little is known about developmental trajectories in the frontoparietal network and there is surprising little consistency in the direction of change, with some studies reporting increased neural activation with age and others decreased neural activation with age (Crone and Dahl, 2012).

An important question in cognitive development concerns the shape of developmental trajectories. One possible hypothesis would be that activity in the frontoparietal network during feedback learning follows a linear trajectory, based on dual-systems models predicting steadily increasing frontoparietal recruitment from childhood to adulthood combined with an adolescent peak in socio-emotional sensitivity in subcortical systems (Ernst et al., 2006; Somerville and Casey, 2010; Steinberg, 2008). On the other hand, prior cross-sectional studies provided preliminary evidence

* Corresponding author at: Institute of Psychology, Brain and Development Lab, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.
E-mail address: s.peters@fsw.leidenuniv.nl (S. Peters).

http://dx.doi.org/10.1016/j.dcn.2016.04.004

for non-linear developmental patterns of frontoparietal activity during feedback learning (Peters et al., 2014a; Van den Bos et al., 2009; Van Duijvenvoorde et al., 2008). These findings indicated that young adolescents are capable of recruiting frontoparietal regions but in different situations than adults, arguing against a simple frontoparietal immaturity model with linear development in cognitive control regions.

Several recent neuroimaging studies have used longitudinal measurements of neural activity to test for neurocognitive changes over development (Ordaz et al., 2013; Paulsen et al., 2015). Longitudinal designs have critical advantages over cross-sectional designs. For instance, previous studies demonstrated important individual differences in developmental trajectories that can be overlooked in cross-sectional designs (Koolschijn et al., 2011; Ordaz et al., 2013; Shaw et al., 2013). Furthermore, longitudinal designs have increased power to detect developmental change, because testing within-individual changes reduces error related to cohort differences (Fjell et al., 2010; Koolschijn et al., 2011). In the current study, neural changes in frontoparietal cortex activity were examined by testing whether frontoparietal activity during feedback learning follows a linear pattern (i.e. monotonic development over time, no adolescent-specific changes), a quadratic pattern (i.e., adolescent-specific effects) or a cubic pattern (adolescent-emergent; e.g. stable levels during childhood, steep changes in adolescence and stabilization in adulthood) (Braams et al., 2015; Somerville et al., 2013). Our longitudinal approach allows for a more specific test of the different hypotheses concerning the pattern of developmental change in frontoparietal areas.

Besides investigating age-related patterns of neural activity, a second goal of this study was to investigate other factors influencing time-related changes in frontoparietal activity in addition to age. There are multiple processes closely related to advancing age that may drive changes in neural activity. That is, an increase in age could be the sole factor explaining time-related increases or decreases in activity, but other factors might also play a role. The factors investigated in this study were task performance, working memory and structural brain development. Task performance has been shown to influence neural activity, and there is evidence that a portion of developmental changes attributed to advancing age are related more to changes in performance (Church et al., 2010; Dumontheil et al., 2010; Koolschijn et al., 2011). Here we tested whether performance on a feedback learning task partly explained changes in neural activation over time. Working memory has previously been argued to be a core prerequisite for cognitive development (Case, 1992) and cognitive control functions (Huizinga et al., 2006), and as such was investigated as an important contributor to changes over time in neural activity during feedback learning. That is, we aimed to study whether a portion of changes in neural activity during feedback learning was explained by individual differences in working memory. A final factor that was investigated is cortical thickness. Several cross-sectional studies have suggested a link between functional activity and structural gray matter in adults (Harms et al., 2013; Hegarty et al., 2012) and children (Dumontheil et al., 2010; Lu et al., 2009; Wendelken et al., 2011). It is likely that developmental changes in neural activity are at least partly influenced by structural development of these brain regions, although the longitudinal relation between structural maturation and development of brain function is not well understood.

Taken together, in this study, we tested developmental trajectories of activation in the frontoparietal network during feedback learning in a large longitudinal fMRI sample across a wide age range (N = 208, 8–27 years) with a two year interval between the first and second time point (see Peters et al., 2014a,b). Our aims were (1) to examine growth trajectories of core areas in the frontoparietal network (DLPFC, SMA, ACC and SPC) and to define the shape of age-related changes, (2) to test the additional contributions of task performance, working memory and structural development to changes over time in neural activity for feedback learning.

2. Methods

2.1. Participants

At time point 1 (TP1), a total of 299 participants between ages 8–27 years underwent an MRI scan, of which 293 participants completed the feedback learning task in the MRI scanner. Of these, 25 participants were excluded from further analyses because of excessive movement (movement >3.0 mm: n = 19), artifacts (n = 3) or because they were extreme outliers in task performance (>3x the interquartile range: n = 3). In total, 268 participants were included at TP1 (*Mean* Age = 14.52 years, *SD* = 3.55; published in Peters et al., 2014a). At time point 2 (TP2), a total of 254 of the initial 299 participants were scanned again approximately two years later (mean time = 1.99 years, SD = 0.10 years, range = 1.66–2.47 years). Reasons for not collecting a scan at TP2 (n = 45) were braces (n = 32) or no interest in participating again (n = 13). Further exclusions at TP2 were because of excessive movement at TP2 (n = 9), scanner artifacts (n = 5), loss of signal (n = 3) or extreme outliers (>3x the interquartile range) on task performance (n = 2).

Only those participants who were included at both TP1 and at TP2 were included in the analyses (N = 208). All analyses were performed on these 208 participants, except for the analyses including working memory and cortical thickness. For working memory, data were incomplete for five participants at TP1 and for two participants at TP2. For the analyses involving structural MRI data, visual quality control led to exclusion of 28 out of 208 participants: Three exclusions for insufficient quality data at both TP1 and TP2, 16 for TP1 and nine for TP2. These participants were only excluded from the analyses where cortical thickness was a factor. Taken together, the analyses with fMRI in the model contained a total of 208 participants (105 females and 103 males), the analyses with working memory a total of 201 participants and the analyses with structural MRI in the model contained a total of 177 participants.

IQ was estimated with two subtests of the WAIS-III or WISC-III (Similarities and Block Design at TP1, Vocabulary and Picture Completion at TP2). The estimated IQ-scores of the 208 included participants were within the normal range at TP1 (85–143, *Mean* = 110.91, *SD* = 9.74) and TP2 (80–147, *Mean* = 108.92, *SD* = 10.18). The study was approved by the Institutional Review Board at the Leiden University Medical Center and all participants (or participants' parents in case of minors) provided written informed consent. Adults received payment for participation and children and their parents received small presents and payment for participation. Participants did not report psychiatric or neurological diagnosis, and no current use of psychotropic medication. All anatomical MRI scans were reviewed and cleared by a radiologist.

2.2. Feedback learning task

Participants performed a child-friendly feedback learning task in the MRI scanner described in detail earlier (Peters et al., 2014a,b). In short, on each trial, participants viewed a screen with three boxes at the top part of the screen (Fig. 1a). At the bottom part of the screen, a stimulus picture was presented, which was one of three possible stimuli. Participants were informed that all pictures belonged in one of the three boxes and that they had to find the correct box for each picture. Performance feedback was provided in the form of a plus-sign ('+') for correct choices (positive feedback) and a minus-sign ('−') for incorrect choices (negative feedback). Stimuli were presented in a pseudorandom order (maximum two identical pictures in a row). The sequence ended after 12

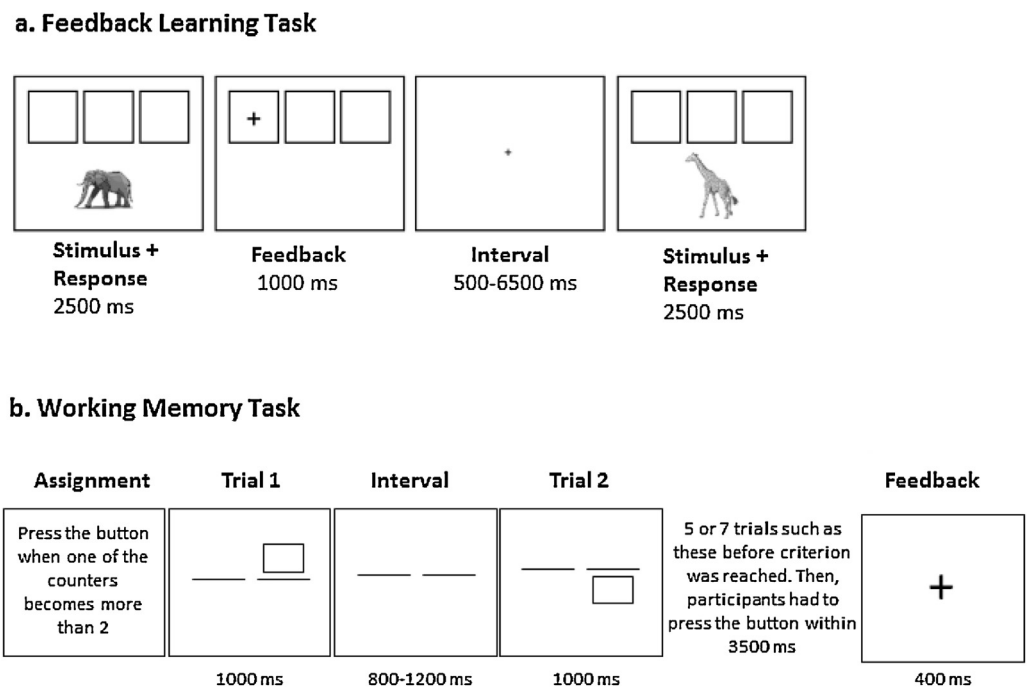

Fig. 1. (a) Display of task sequence for the feedback learning task. During the last 500 ms of the Interval screen, a fixation cross was presented to prepare the participant for the next upcoming stimulus. (b) Display of task sequence for the Mental Counters task as a measure of working memory.

trials, or when participants chose the correct location twice for all three stimuli. Subsequently, a new sequence with three new pictures was presented. In total, participants completed 15 sequences, which resulted in a maximum of 180 trials. On average, participants needed 138 trials (average 9.2 per sequence) to complete the task at T1 and 136 trials at T2 (average 9.1 per sequence). The task was divided into two blocks of eight and seven sequences, respectively. Before performing the task in the MRI scanner, participants practiced three sequences in a separate practice session. All trials started with a 500 ms fixation cross, followed by a 2500 ms time window during which the stimulus was presented and a response needed to be given. Feedback was presented for 1000 ms. Inter-trial intervals were jittered with OptSeq (Dale, 1999), with intervals between 0 and 6 s in addition to the 500 ms fixation cross.

2.3. Feedback types

We distinguished between a learning phase and an application phase for all stimuli. The learning phase was defined as the trials where participants had not yet found the correct location for each stimulus and were still using feedback to find the correct locations. The application phase was defined as the trials where each stimulus was sorted correctly previously and was continued to be sorted correctly on subsequent trials. We excluded trials in the learning phase that did not result in learning, i.e., the trials where the feedback was not successfully used on the subsequent trial (5.71% of the trials). Based on the learning-application distinction, we defined the following three feedback types: Learning phase. (a) Positive Learning: A first correct feedback for a stimulus followed by a correct sort on the next trial for this stimulus. (b) Negative Learning: A first encountered incorrect feedback for a stimulus followed by a choice for another location on the next trial of this stimulus. Application phase. (c) Application: Correct (i.e., positive) feedback for a stimulus that was sorted correctly before.

As a task performance measure, we calculated the 'learning rate' for each participant. This was defined as the percentage of trials in the learning phase for which feedback was successfully used on the next trial, compared to the total number of trials during the learning phase (including trials which did not result in learning according to the participants' behavior on the next trial).

2.4. Working memory task

Working memory was assessed outside of the MRI scanner with the Mental Counters task (Larson et al., 1988), which has been shown in a prior developmental study to be a well-suited task to measure the latent factor of working memory in children (Huizinga et al., 2006). Similar to the feedback learning paradigm, the mental counters task has both a spatial aspect (because two counters at two different locations have to be remembered) and a verbal rehearsal component, allowing us to assess the working memory aspects of the feedback learning task contributing to neural activity. Participants were presented with a screen with two horizontal lines (the 'counters') placed next to each other (Fig. 1b). At each trial, a square randomly appeared on top of or below one of the two horizontal lines. The participant was instructed to keep track of the 'score' of the two counters. The value of the counters changed on each trial: e.g., a square appearing above the left counter changed the score to 1-0. If on a next trial a square appeared above the right counter, the score changed to 1-1. A square appearing below one of the counters meant a point had to be subtracted for that counter. Participants were instructed to press a button when one of the counters reached a certain criterion value, e.g. 'press when the score for one of the counters reaches more than two points', and the criterion changed for each series. The amount of trials before criterion was reached was set to either five or seven trials. In total, 16 series were presented. Trials were separated by 800–1200 ms intervals and participants had a time window of 3500 ms to respond once criterion was reached. Feedback was provided as a '+' for a correct button press, a '–' for an incorrect button press and an "x" for an omission. Performance on the working memory task was defined as the proportion of correct responses (at TP1: $M = 0.83$, $SD = 0.15$; at TP2: $M = 0.87$, $SD.12{=}0.12$).

2.5. FMRI data acquisition

We used the same Philips 3.0 T MRI scanner and settings for both time-points (Peters et al., 2014a,b). Functional scans were acquired with T2*-weighted echo-planar imaging, for which the first two volumes were discarded to allow for equilibration of T1 saturation effects. The following settings were used: TR = 2.2 s, TE = 30 ms, sequential acquisition, 38 slices, slice thickness = 2.75 mm, Field of View (FOV) = 220 × 220 × 114.68 mm. We acquired a high-resolution 3D T1-FFE anatomical scan after the experimental task (TR = 9.76 ms, TE = 4.59 ms, 140 slices, voxel size = 0.875 × 0.875 × 1.2 mm, FOV = 224 × 177 × 168 mm). The experimental task was projected on a screen that was viewed through a mirror attached to the head coil. Participants were accustomed to the MRI environment and sounds with a mock scanner prior to the actual MRI scan.

2.6. FMRI data analysis

We performed two types of analyses: a whole-brain analysis for an illustrative overview of brain activity at TP1 and TP2, and regions-of interest (ROI) analyses for growth curve modeling. For all analyses we used SPM8 (Wellcome Department of Cognitive Neurology, London) to analyze fMRI data. All scans were corrected for slice timing acquisition and rigid body motion. All volumes were spatially normalized to T1 templates, using a 12-parameter affine transform with a nonlinear transformation involving cosine basis functions with resampling of the volumes to 3 mm voxels. T1 templates were based on the MNI305 stereotaxic space (Cocosco et al., 1997), an approximation of Talairach space (Talairach and Tourneaux, 1988). Functional volumes were spatially smoothed with an 8 mm FWHM isotropic Gaussian kernel. The fMRI time series data were modeled by convolving a series of events with a hemodynamic response function. The modeled feedback events were categorized as: "Positive Learning", "Negative Learning", and "Application", which were time-locked with 0-duration to the moment of feedback presentation. All other trials (e.g., trials that did not result in learning or too-late trials) were modeled as events of no interest. These events were used as covariates in a general linear model together with a set of cosine functions that high-pass filtered the data. The least-squares parameter estimates of height of the best-fitting canonical HRF for each condition were used in pair-wise contrasts.

The main fMRI contrast was the learning contrast: Learning (Positive and Negative combined) > Application. With this contrast, we investigated neural activity for feedback that is informative for learning, averaged across negative and positive valence, relative to application. In a prior cross-sectional paper (Peters et al., 2014a) we focused more specifically on valence differences across development. The contrast images were submitted to higher-level group analyses. Whole-brain fMRI analyses were performed with an FWE-corrected threshold at $p < 0.05$.

2.7. Region-of-Interest analyses

Region-of-interest (ROI) analyses were performed with the MarsBaR toolbox (v. 0.42) in SPM8 (Brett, Anton, Valabregue, & Poline, 2002). We used bilateral anatomical ROIs which were obtained from the probabilistic Harvard-Oxford Cortical Structural atlas. The ROIs were Middle Frontal Gyrus for DLPFC, Superior Parietal Lobule for SPC, Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex) for SMA, and Cingulate Gyrus (anterior division) for ACC. These regions were selected based on prior studies that revealed age effects for feedback learning in these regions (Peters et al., 2014a; Van Duijvenvoorde et al., 2008). The anatomical region for SMA in addition to the ACC was chosen because the SMA region better captures the peak voxels of neural activity related to performance monitoring as described in a meta-analysis by Ridderinkhof et al. (2004). In addition, a review of feedback processing regions by Crone (2014) also referred to this same anatomical region which is more dorsal than the anatomical ACC region. Finally, in a prior study describing cross-sectional age comparisons in this sample (Peters et al., 2014a,b), the peak voxels that showed an age-effect in the Learning > Application contrast were located in the anatomical SMA region and not in the ACC region. To get a complete picture of contributions from medial prefrontal cortex to feedback learning, we have also added a ROI based on the anatomical region for ACC, resulting in a total of four ROIs (DLPFC, SPC, SMA and ACC).

Because the anatomical ROIs based on probability maps were large and to ensure the selected voxels had a high probability of belonging to the targeted ROI, we created more focal ROIs by thresholding at 50%, indicating that for each voxel, the probability that the voxel was actually part of the ROI (e.g. DLPFC) was >50% (see Fig. 2b). A visual comparison confirmed that these anatomical ROIs overlapped with the activity clusters for the contrast Learning > Application (FWE-corrected at $p < 0.05$) (see Fig. 2a). Beta values reflecting activity for all voxels within each ROI were averaged to produce a mean signal for each ROI per time point.

2.8. Structural brain analysis

Cortical reconstruction was measured automatically using FreeSurfer 5.3 (http://surfer.nmr.mgh.harvard.edu). Structural brain maturation of TP1 was reported earlier using a prior version of FreeSurfer (Koolschijn et al., 2014), therefore we reconstructed and reanalyzed all anatomical scans from the first wave using version 5.3. Details of the surface-based cortical reconstruction and subcortical volumetric segmentation procedures have been extensively documented previously (Dale et al., 1999; Fischl and Dale, 2000; Fischl et al., 2004; Ségonne et al., 2004). To extract reliable volume and thickness estimates, images where automatically processed with the longitudinal stream in FreeSurfer (Reuter et al., 2012). Specifically, an unbiased within-subject template space and image (Reuter and Fischl, 2011) is created using robust, inverse consistent registration (Reuter et al., 2012). Several processing steps, such as skull stripping, Talairach transforms, atlas registration as well as spherical surface maps and parcellations were then initialized with common information from the within-subject template, significantly increasing reliability and statistical power (Reuter et al., 2012). To extract average cortical thickness from the FSL anatomical ROIs, we performed the following steps: (1) Each anatomical ROI (DLPFC, SMA, ACC and SPC) was registered automatically to the FreeSurfer "fsaverage" template with normalized mutual information and inspected for accuracy of registration. Of note, as FreeSurfer calculates cortical thickness per hemisphere, the ROIs were split into a left and right structural ROI. (2) Individual cortical thickness data was mapped to the "fsaverage" template. (3) Average cortical thickness in mm was extracted for each ROI and individual separately.

2.9. Statistical analyses

To model the shape of individual growth curves, we used mixed model analyses (also termed "random effects", "multilevel modeling", or hierarchical linear model-analyses) on the ROI values for contrasts of interest (Ordaz et al., 2013). This method expands on multiple regression analyses and is suited for longitudinal data because it takes into account the repeated-nature of the data, and controls for the dependency in measures within individuals (i.e., nested data). Change scores were not calculated, because mixed models take into account all data including individual differences

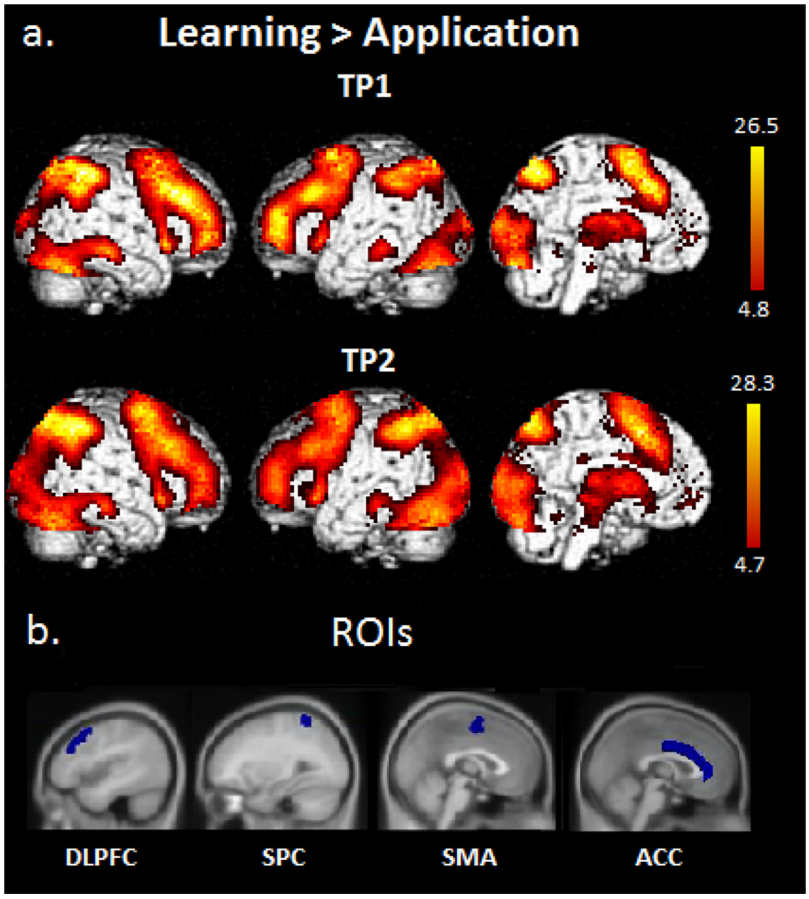

Fig. 2. (a) Whole-brain analyses showing comparable neural activation patterns at TP1 and TP2 (FWE-corrected at $p < 0.05$). (b) Bilateral Regions of Interest in the DLPFC, SPC, SMA and ACC, extracted from the Harvard-Oxford Cortical Atlas. The anatomical ROIs were situated within the activation maps of the Learning > Application contrast.

in intercepts. We performed mixed analyses with the NLME package in R (Pinheiro et al., 2007) version 3.1.0. With this package it is possible to test for fixed effects (effects that are similar for all participants) and random effects (effects that vary across participants) of age on brain activity. Models were compared using the Akaike Information Criterion (AIC), a standard measure for model comparison which indicates how well the model describes the data. Lower AIC values indicate a better fit of the model to the data. For nested models (i.e., comparing models with only 1 different term), we additionally tested with log-likelihood tests (χ^2) whether changes in model fit were significant. Our goals were twofold: (1) to test which shape of age (linear, quadratic cubic) best described the developmental pattern for the variables neural activity, task performance, working memory and cortical thickness, and (2) to investigate which factors explain variance in brain activity above age; task performance, working memory and/or cortical thickness. The model-building steps are described in the next paragraphs.

2.9.1. Developmental patterns: linear, quadratic and cubic trajectories

We first tested for each dependent variable (task performance, working memory and activity and cortical thickness for each ROI) which age shape best described the developmental pattern (phase (1). In Table 1, we provide an overview of the model-building steps. The base-model included a fixed intercept and a random intercept, with the latter capturing the variation in the intercept to account for the repeated nature of the data. Next, the base-model was tested against three models that tested the shape of the grand mean trajectory for age. We tested for a linear effect of age (i.e., monotonic development), a quadratic effect of age (i.e., an adolescent-specific effect) and a cubic effect (i.e., adolescent-emergent pattern) by adding three polynomial functions for age to the base-model (Braams et al., 2015; Somerville et al., 2013). We only selected a linear, quadratic or cubic model if the age term resulted in a better fit compared to the base-model without age as indicated by the AIC and a log-likelihood test. For the best model, we tested whether specifying age as an effect with a random slope resulted in a better fit (judged by the AIC and a log-likelihood test) compared to age as a fixed effect. A significant random slope would indicate that the effect of age on the dependent measure differs for each individual.

An example of a formal notation of such a mixed model to predict, for instance, task performance, would be as follows:

$$\text{Performance} = \pi_{0i} + \pi_{1i}(\text{Age})_{ti} + e_{ti}$$

with

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

In this example model, substitution of the second level model into the first level model gives the integrated model that was fitted to the data. β_{00} reflects the grand mean intercept of neural activity at the average age of the sample. β_{10} reflects the grand mean slope of age effects on neural activity, and e_{ti} represents the residual error term. This example model displays a random intercept (r_{0i}) indicating different starting points of development for each

Table 1
Model building steps for developmental trajectories of neural activity (phase one) and for individual differences influencing neural activity in addition to age (performance, working memory and cortical thickness; phase two).

Model-building steps	
Phase 1: Developmental trajectories	
1	Base model (fixed & random intercept only)
2	Basemodel + Age
3	Basemodel + Age + Age^2
4	Basemodel + Age + Age^2 + Age^3
Phase 2: Additional effects of individual differences over age	
1	Best model from Phase 1
2	Best model from Phase 1 + task performance
3	Best model from Phase 1 (+ task performance if significant) + working memory
	Best model from Phase 1 (+ task performance if significant) (+ working memory if significant) + cortical thickness

Models were compared using the Akaike Information Criterion (AIC). For nested models (i.e., comparing models with only 1 different term), we additionally tested with log-likelihood tests (χ^2) whether changes in model fit were significant.

participant, and a random slope of age (r_{1i}), indicating individual variability in the change over time. All models were fit with full information maximum likelihood estimates. A random slope of age did not improve model fit in any of the analyses we performed, except for the quadratic age effect on working memory ($p < 0.001$). Therefore, all models reported in the results section are without a random effect of age.

2.9.2. Explaining development of neural activity with different predictors

In the second phase, we tested which factors explain development of neural activity in the frontoparietal network in addition to age. The predictors we tested for significance above age were task performance, working memory and cortical thickness. Therefore, we tested a combined model including all predictors to account for neural activity in the four ROIs. The model-building procedure consisted of multiple steps. We started with the best fitting age-model (linear, quadratic or cubic) for the ROI determined in the previous analysis. To test wether other measurements explained additional variance above age, we first added task performance as a second predictor to investigate whether performance explained additional variance above age. If this was the case, performance was included in the next step and if this was not the case, we continued to the next step with only the linear, quadratic or cubic age term. The next step was adding working memory the model, followed by cortical thickness of the ROI as a final step. AIC and log-likelihood values were used to test whether model fit improved by adding the predictor. Changing the order of adding performance, working memory and cortical thickness did not change the overall results.

2.10. Reliability

For visualization purposes, the whole-brain results for the Learning > Application contrast for TP1 and TP2 across all participants are displayed in Fig. 2a and Table 2. The results showed that a comparable network was recruited on TP1 and TP2, including bilateral DLPFC, SMA, ACC and bilateral SPC. For all predictors (ROI activity and cortical thickness, task performance and working memory), we also assessed reliability in ROIs from TP1 to TP2. We calculated intra-class correlation coefficients (ICCs) on the mean signal of each ROI. We used a two-way mixed model with absolute agreement and we reported the average measure. A value of 0 indicates no relation between the first and second time point and a value of 1 indicates perfect agreement. Interpretation of ICC values for reliability was guided by Cicchetti (2001): values<0.4 were interpreted as poor; values 0.41-0.59 were interpreted as fair, values 0.60-0.74 were interpreted as good, and values >0.75 were interpreted as excellent. As can be seen in Fig. 3, most ICC values were in the 'fair to good' range. For a visual comparison of

Table 2
MNI coordinates of local maxima activated for the contrast Learning > Application at T1 and T2 across all participants, FWE-corrected at $p.05<0.05$. Subpeaks for larger clusters were reported to a maximum of 10 with the highest T value (which were >4.0 mm apart from each other).

Area of activation	x	y	z	voxels	T
TIMEPOINT 1					
R middle frontal gyrus	45	29	37	9927	24.05
R superior medial gyrus	3	23	43	s.c.	23.66
R supplementary motor area	6	20	46	s.c.	23.59
R superior frontal gyrus	24	8	61	s.c.	23.07
L superior medial gyrus	−3	26	40	s.c.	22.00
R insula lobe	33	23	1	s.c.	20.22
L middle frontal gyrus	−45	20	37	s.c.	19.44
R middle frontal gyrus	33	56	4	s.c.	19.03
L middle frontal gyrus	−33	50	22	s.c.	18.71
L insula lobe	−30	20	4	s.c.	18.68
R inferior parietal lobule	48	−43	52	8246	23.01
R inferior parietal lobule	39	−43	43	s.c.	21.30
R inferior parietal lobule	39	−52	52	s.c.	21.14
R inferior parietal lobule	39	−52	46	s.c.	21.10
R inferior parietal lobule	45	−52	46	s.c.	20.93
L inferior parietal lobule	−51	−40	52	s.c.	19.76
R angular gyrus	33	−58	49	s.c.	19.40
R cerebellum	36	−61	−32	s.c.	18.77
L inferior parietal lobule	−48	−49	49	s.c.	18.45
R superior parietal lobule	33	−64	52	s.c.	18.37
L middle temporal gyrus	−57	−31	−8	151	10.40
TIMEPOINT 2					
R inferior parietal lobule	45	−43	49	120510	28.47
R inferior parietal lobule	39	−49	49	s.c.	24.86
R inferior parietal lobule	36	−49	43	s.c.	24.37
R angular gyrus	39	−55	52	s.c.	23.93
R angular gyrus	33	−61	49	s.c.	23.40
L inferior parietal lobule	−45	−43	46	s.c.	22.65
R precuneus	6	−70	58	s.c.	22.52
R inferior parietal lobule	48	−52	46	s.c.	22.47
L inferior parietal lobule	−33	−43	40	s.c.	21.85
L superior parietal lobule	−27	−64	46	s.c.	21.58
R supplementary motor area	3	17	49	118160	26.61
L supplementary motor area	−3	17	49	s.c.	25.92
R middle frontal gyrus	30	8	55	s.c.	25.30
R insula lobe	30	23	−2	s.c.	25.28
R superior frontal gyrus	27	2	55	s.c.	25.17
R superior frontal gyrus	30	2	61	s.c.	24.17
L insula lobe	−30	23	1	s.c.	22.97
R middle frontal gyrus	45	26	37	s.c.	21.07
L middle frontal gyrus	−27	2	58	s.c.	20.08
R middle cingulate cortex	6	26	34	s.c.	19.48
no label	0	−31	28	26	6.84
Cerebellar vermis	0	−49	−17	18	6.10

Abbreviations: L = Left, R = Right, s.c. = same cluster.

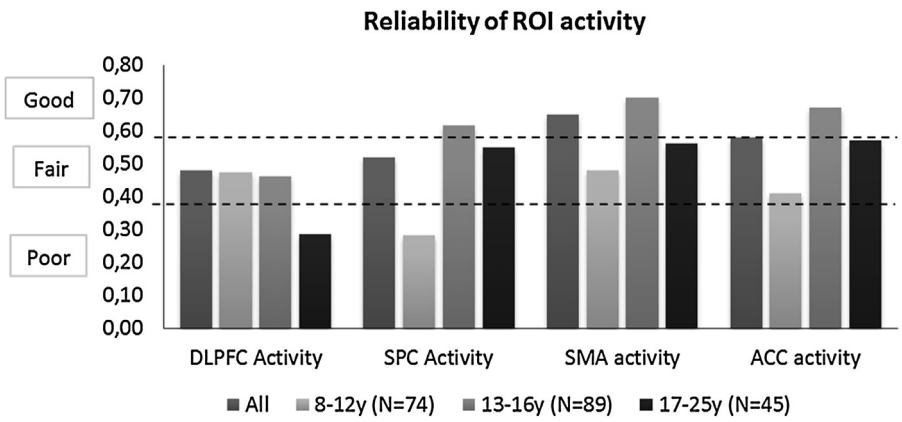

Fig. 3. Intra-Class Correlations (ICC) values for each ROI per age group. The labels 'poor', 'fair', 'good' and 'excellent' are based on Cicchetti (2001).

ICCs across different age groups, Fig. 3 also shows ICC values for 3 age groups (8–12 years, 13–16 years, 17–25 years). For DLPFC, reliability was lowest (in the poor range) in the oldest age groups (suggesting change over time), and was fair for the children and adolescents. In contrast, for SPC, SMA and ACC, reliability was lowest in the child group (suggesting change over time), and was fair to good for adolescents and adults.

3. Results

3.1. Developmental trajectories: behavioral measures

Behavioral results indicated that all participants performed adequately on the feedback learning task and the working memory task. Learning rate (i.e., the percentage of trials during the learning phase, for which feedback was successfully used in a subsequent trial) was relatively high at both T1 (M = 93.82, SD = 4.76, range = 71–100) and T2 (M = 94.86, SD = 4.59, range = 79–100). Performance on the working memory task (performed outside of the scanner) was also adequate at both T1 (M=0.83, SD=0.15, range=0.13–1.00) and T2 (M=0.87, SD=0.12, range=0.31–1.00). For descriptive purposes, correlations between all measures (feedback learning performance, working memory, neural activity and cortical thickness) are presented in Table 3.

Next, we started the model building procedure by testing for both measures (task performance and working memory) whether a linear, quadratic or cubic age pattern best described developmental change. AIC values and a log-likelihood test were used to test which model best fit the data. AIC values for the base model (without age), linear, quadratic and cubic age model for each measure are listed in Table 4.

For task performance on the feedback learning task (learning rate: the percentage of feedback that was successfully applied in a subsequent trial), there was a linear and quadratic effect of age on task performance. Note that although the AIC value was slightly lower for a cubic model than for a quadratic model, this difference was not significant according to a log-likelihood test (p.084=0.084). This indicated that task performance improved with age, and then leveled off for older participants (see Table 4 and Fig. 4 for the raw data and predicted data).

The age models for working memory indicated that a combined linear and quadratic model described the best fit. Note that the AIC value was slightly lower for the cubic age model, but the accompanying log-likelihood test showed no significant improvement (p = 0.58) (see Table 4; Fig. 4). This indicated that working memory improved with age, and then leveled off for older participants.

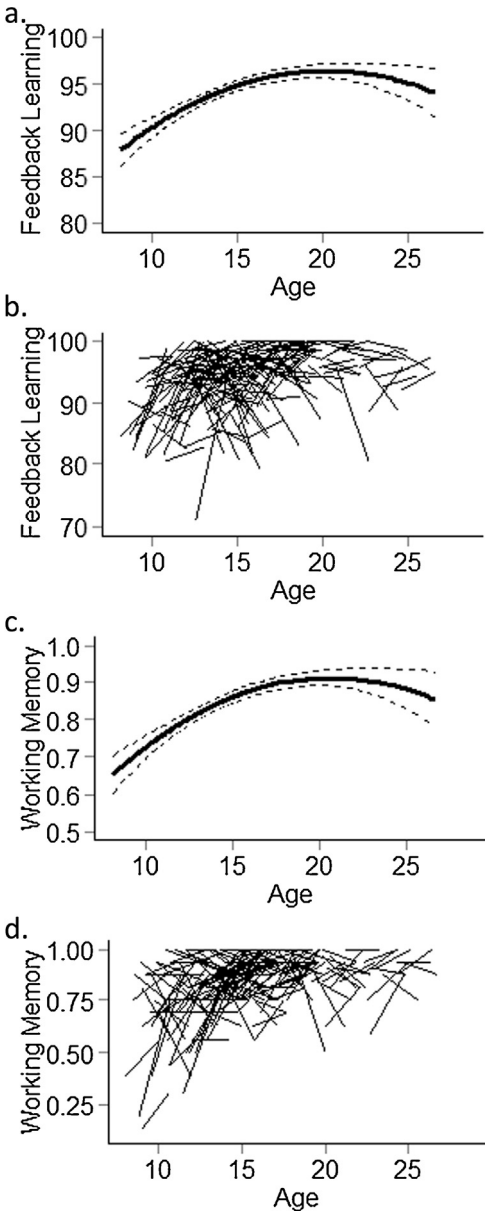

Fig. 4. Predicted data (a and c) and raw data (b and d) for feedback learning performance and working memory. Dotted lines represent 95% confidence intervals. Feedback learning performance was defined as the percentage of trials during the learning phase where (positive or negative) feedback was successfully used on a subsequent trial.

Table 3
Correlations between all variables at both time points. **: Significant at $p < 0.01$, *: Significant at $p < 0.05$. Abbreviations: CT = cortical thickness, WM = working memory, Perf = performance, n.s. = not significant.

Time point	DLPFC	SPC	SMA	ACC	DLPFC CT	SPC CT	SMA CT	ACC CT	WM	Perf	Age TP
Time point 1											
DLPFC	1										
SPC	0.463**	1									
SMA	0.489**	0.574**	1								
ACC	−0.390**	0.633**	0.178*	1							
DLPFC CT	−0.192**	*n.s.*	*n.s.*	*n.s.*	1						
SPC CT	*n.s.*	*n.s.*	*n.s.*	*n.s.*	0.390**	1					
SMA CT	*n.s.*	*n.s.*	*n.s.*	*n.s.*	0.423**	0.332**	1				
ACC CT	*−0.287***	*n.s.*	*n.s.*	*0.170**	0.341**	0.178*	0.306**	1			
WM	0.203**	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	−0.144*	−0.227**	1		
Perf	0.343**	0.222**	0.155*	*n.s.*	−0.140*	*n.s.*	*n.s.*	*−0.244***	.437**	1	
Age TP1	0.410**	0.205**	0.210**	−0.145*	−0.382**	−0.162*	−0.307**	−0.546**	.340**	.444**	1
Time point 2											
DLPFC	1										
SPC	0.326**	1									
SMA	0.400**	0.389**	1								
ACC	−0.153*	0.335**	*n.s.*	1							
DLPFC CT	−0.146*	*n.s.*	−0.170*	*n.s.*	1						
SPC CT	*n.s.*	*n.s.*	*n.s.*	*n.s.*	0.383**	1					
SMA CT	*n.s.*	*n.s.*	*n.s.*	*n.s.*	0.491**	0.341**	1				
ACC CT	*n.s.*	*n.s.*	*n.s.*	*0.179***	0.340**	*n.s.*	0.320**	1			
WM	0.165*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*−0.199***	1		
Perf	0.227**	*n.s.*	0.159*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	0.285**	1	
Age TP2	0.209**	*n.s.*	0.296**	−0.144*	−0.380**	*n.s.*	−0.307**	−0.503**	0.311**	0.167*	1

Table 4
AIC and loglikelihood *p*-values for the base model, and linear, quadratic and cubic age models for brain activity, cortical thickness, task performance and working memory. The best fitting model is highlighted with a bold font. Abbreviations: CT = cortical thickness, AIC = Akaike Information Criterion.

	Base	Linear		Quadratic		Cubic	
	AIC	AIC	*p*	AIC	*p*	AIC	*p*
DLPFC activity	1562.10	1537.33	<0.001	**1533.26**	0.014	1535.09	0.675
SMA activity	1872.18	**1851.03**	<0.001	1851.70	0.249	1853.45	0.618
SPC activity	1688.66	1690.18	0.486	**1683.57**	0.003	1685.30	0.600
ACC activity	1679.57	**1668.97**	<0.001	1670.41	0.457	1671.48	0.334
DLPFC CT	−425.54	**−553.83**	<0.001	−553.17	0.248	−553.715	0.110
SMA CT	−336.29	−421.659	<0.001	−421.60	0.164	**−430.81**	<0.001
SPC CT	−447.86	−478.99	<0.001	**−488.05**	<0.001	−486.25	0.656
ACC CT	−550.41	−732.47	<0.001	**−747.36**	<0.001	−748.81	0.063
Performance	2442.43	2406.12	<0.001	**2387.45**	<0.001	2386.47	0.084
Working Memory	−485.96	−529.56	<0.001	**−546.92**	<0.001	−548.52	0.057

3.2. Developmental trajectories: neural measures

The same model building procedure was used for testing the best fitting shapes of the neural measures (neural activity and cortical thickness). For neural activity (i.e., the Learning > Application contrast), we observed distinct developmental changes across the different ROIs. See Table 4 for the model comparison values and Fig. 5 for the predicted and actual data for each ROI. For DLPFC, the relationship between age and neural activity was best described by both a linear and a quadratic term for age, i.e. activity increased with age, and leveled off in late adolescence and young adulthood. For SPC, there was a quadratic but not a linear effect of age, which indicated that activity increased until adolescence and then decreased into adulthood. In the SMA, neural activity was best described by a linear effect of age that showed increasing neural activity with increasing age. In contrast, neural activity in ACC showed a linear decrease with age. In all ROIs, there was significant individual variability in mean neural activation as indicated by a random effect of the intercept.

For cortical thickness of the ROIs, we found different developmental patterns depending on the region. For DLPFC, a linear model with decreasing cortical thickness with age best described the data. In contrast, in SPC and ACC the relationship between age and cortical thickness was best described by a quadratic model with a significant linear and quadratic term for age. That is, cortical thickness decreased with age and stabilized in late adolescence/young adulthood (see Fig. 6). For SMA, a model with both a linear and cubic (but not quadratic) effect of age best described the data, i.e., cortical thickness showed a relatively stable pattern in young adolescents, decreased steeply in adolescence and stabilized in late adolescence/early adulthood (see Fig. 6). All results are described in Table 4.

3.3. Explaining change with age, performance, working memory and cortical thickness

To investigate which factors additionally covaried with time-related changes in neural activity, we tested the contributions of task performance, working memory, and cortical thickness in addition to age in a hierarchical mixed model. Neural activity (i.e., the Learning > Application contrast) was the dependent variable (for each ROI separately) and the predictors task performance, working memory and cortical thickness were added in a consecutive order above age. The starting model was the model with the best fitting age shape (linear, quadratic or cubic). Next, task performance, working memory and cortical thickness were added in hierarchical steps (see Methods section). The final model parameters for the best fitting model are presented in Table 5.

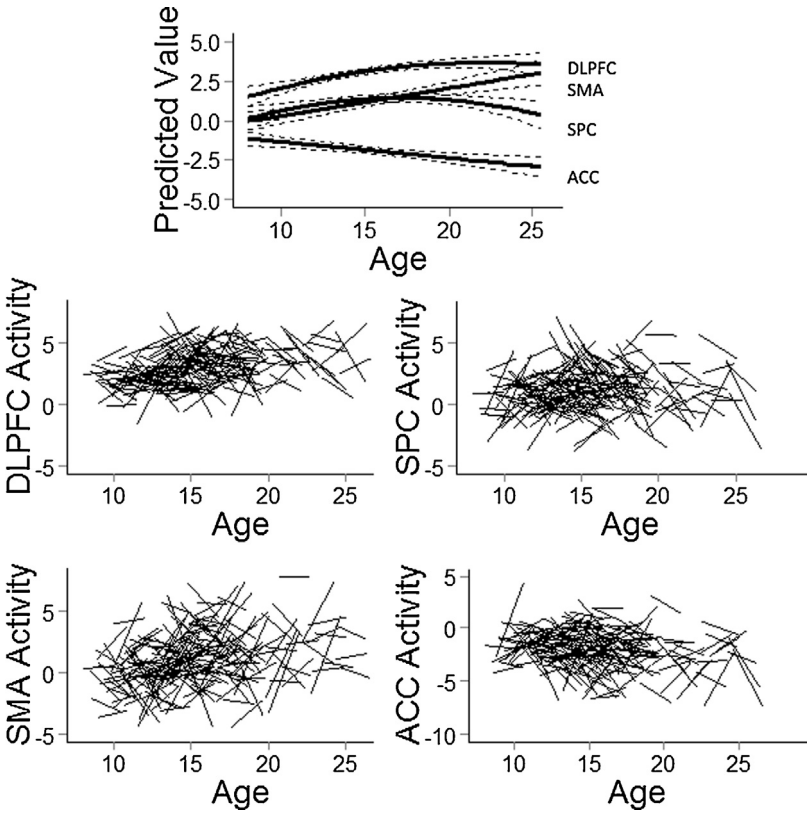

Fig. 5. Predicted data (presented in the upper graph) and raw data for neural activity in DLPFC, SPC, SMA and ACC. Dotted lines represent 95% confidence intervals. Each line represents one individual at two time points.

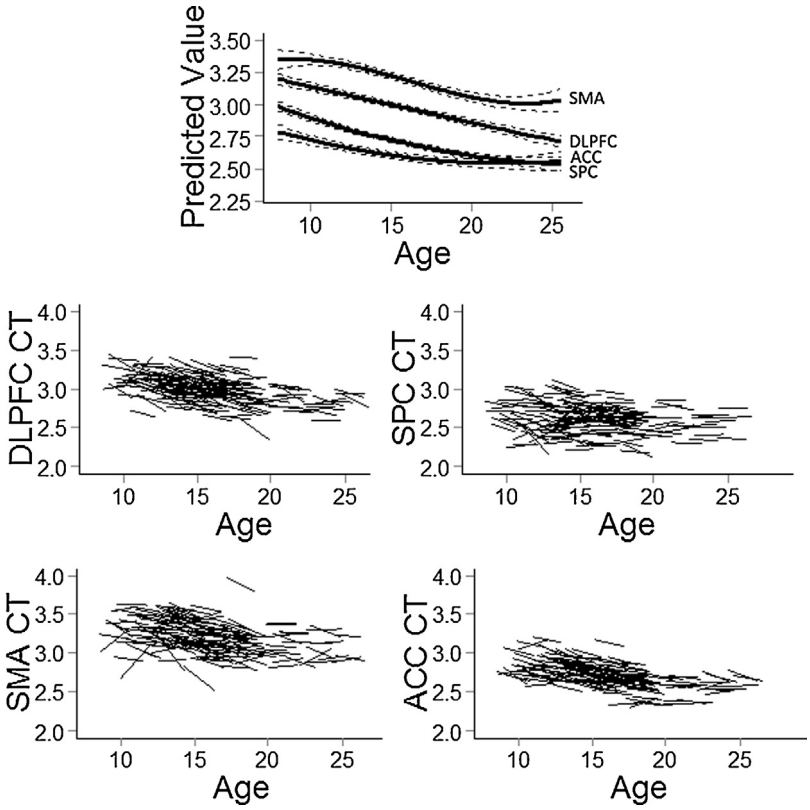

Fig. 6. Predicted data (presented in the upper graph) and raw data for cortical thickness in DLPFC, SPC, SMA and ACC. Dotted lines represent 95% confidence intervals. Each line represents one individual at two time points.

Table 5
Model parameters for the best fitting model for DLPFC, SPC, SMA and ACC. CT = cortical thickness, CI = confidence interval.

Area	Variance	β	P	95% CI	
				Lower	Upper
DLPFC					
Random effect					
Intercept	0.69			0.49	0.98
Fixed effects					
Intercept		3.09	<0.001	2.92	3.25
Age		5.89	<0.001	2.57	9.22
Age^2		−2.17	0.18	−5.35	1.00
Performance		0.07	<0.001	0.035	0.11
SPC					
Random effect					
Intercept	1.09			0.087	1.36
Fixed effects					
Intercept		1.18	<0.001	0.96	1.39
Age		−0.46	0.833	−4.74	3.92
Age^2		−4.36	0.036	−8.4	−0.31
Performance		0.05	0.044	0.0015	0.09
SMA					
Random effect					
Intercept	1.51			1.26	1.82
Fixed effects					
Intercept		1.29	<0.001	1.01	1.56
Age		0.20	<0.001	0.12	0.28
CT		1.48	0.042	0.062	2.85
ACC					
Random effect					
Intercept	1.20			1.00	1.45
Fixed effects					
Intercept		−1.98	<0.001	−2.20	−1.76
Age		−0.010	<0.001	−0.016	−0.043

For DLPFC, the model that best explained neural activity was a model including a linear (but no longer quadratic) age effect and a positive effect of task performance ($p < 0.001$), i.e. better performance predicted increased activity (Table 5). Working memory (p=0.724) and cortical thickness (p=0.066) did not explain additional variance over and above age and task performance. For SPC, task performance explained a significant amount of variance in neural activity (p=0.044), such that in addition to a quadratic age effect, better task performance was associated with increased activity (Table 5). Similar to the DLPFC, working memory (p=0.426) and cortical thickness (p=0.913) did not contribute to the model over and above age and task performance. A different pattern was found for SMA. Here, we observed a significant positive linear effect for age and cortical thickness (p=0.042; Table 5), but not for task performance (p=0.081) and working memory (p=0.623). The model indicated a positive relation between cortical thickness and neural activity, such that increased activity was associated with increased cortical thickness. For ACC, neither task performance (p=0.275), working memory (p=0.190) or cortical thickness (p=0.200) explained additional variance above age (Table 5).

4. Discussion

The main aim of this study was to examine the developmental trajectory of neural activity in the frontoparietal network during a feedback learning task. We tested for different developmental trajectories (linear, quadratic and cubic) in neural activation and for factors contributing to time-related changes in brain activity for feedback learning above age, particularly task performance, working memory and cortical thickness as an index of structural brain development. The results showed differential developmental trajectories for the key brain regions involved in learning from feedback (DLPFC, SPC, SMA and ACC). These findings are discussed in more detail in the next paragraphs.

4.1. Growth trajectories for neural activity, cortical thickness, performance and working memory

Mixed model analyses of the longitudinal data showed that age-related changes in neural activity followed different developmental trajectories for key regions in the frontoparietal network. The linearly increasing pattern in SMA activity fits with prevailing developmental theories of monotonously increasing cognitive control (Ernst et al., 2006; Somerville and Casey, 2010). Interestingly, ACC activity showed a linearly decreasing pattern of activity with age. Possibly, this indicates a shift from reliance on ACC to SMA when learning from feedback across adolescence, or more specialization with increasing age (Johnson, 2011).

Intriguingly, the current findings show peak activity in SPC and to a lesser extent in DLPFC during late adolescence when learning from feedback. For SPC, there was a quadratic pattern with a peak in mid/late-adolescence, whereas for DLPFC, there was a combined linear and quadratic pattern with peaking and leveling off in late adolescence/early adulthood. Previous studies found contradicting patterns for frontoparietal recruitment, with some showing increases in activation with increasing age and others showing decreases, and both patterns were interpreted as reflecting immaturity in adolescence (Pfeifer and Allen, 2012). The quadratic pattern in DLPFC may seem to contradict initial theories of prefrontal cortex maturation, which suggested a linearly protracted developmental pattern until the early twenties. A closer inspection of the studies reported to date shows that no prior large-scale study included our age range of 8–27 years. Instead, prior studies selected age groups, such as adolescents aged 13–17 versus adults aged 18–25 years or other age selections that did not span the whole range of adolescence and early adulthood (Geier et al., 2009; Thomason et al., 2009; Van den Bos et al., 2009; Van Duijvenvoorde et al., 2008; Velanova et al., 2008). This is problematic, for instance given that participants in 'adult' groups continue to develop in terms of cognitive control and neural activity at least until the early twenties (Cohen et al., 2016). Thus, these prior studies may have been underpowered to detect a quadratic pattern, given that most prior studies used cross-sectional comparisons (Fjell et al., 2010).

One possible interpretation of the quadratic patterns in DLPFC and SPC activity may be in terms of increased potential for learning and flexibility in late adolescence. Recent research showed that complex paradigms such as divergent thinking resulted in stronger DLPFC activity in adolescents than in adults (Kleibeuker et al., 2013). Studies using more basic cognitive control tasks, however, have not reported these peak activations in frontoparietal regions in adolescence (Klingberg et al., 2002; Rubia et al., 2006). Therefore an important direction for future research is to unravel whether, when, and how cognitive performance and DLPFC and SPC show quadratic effects in late adolescence. Recently, it was found that adolescents recruit DLPFC more strongly than adults when financial incentives were offered for performing well (Teslovich et al., 2014), suggesting that adolescents may engage DLPFC more in a context of high motivation.

4.2. Additional predictors of frontoparietal activity over age

We also examined the effects of individual differences on time-related changes in neural activity besides age, such as performance, working memory and cortical thickness. We first tested the general age patterns for these variables. For cortical thickness, all areas showed a decrease with age, which fits with prior studies showing an initial increase in childhood followed by cortical thinning in adolescence (Koolschijn and Crone, 2013; Shaw et al., 2013; Tamnes

et al., 2010). Both performance on the feedback learning task and working memory showed a steadily increasing performance in adolescence which leveled off in adulthood (quadratic pattern). A close examination of the literature shows that few studies examined cognitive control development across the whole age range of adolescence until adulthood. For example, a large and comprehensive study on the development of executive functions compared children of 7, 11, 15 and 21 years (Huizinga et al., 2006), but not the intermediate ages. Likewise, a study on performance monitoring previously compared performance of 8–9 year-olds, 11–13-year-olds and 18–25-year-old adults, but did not separate between ages within the adult group (Van Duijvenvoorde et al., 2008).

An important question is whether changes over time in neural activity are related to age per se or also to performance changes (Church et al., 2010). Therefore, we tested if performance explained additional variance above age the regions of interest. This was the case for DLPFC and SPC, with better performance linked to increased activity. This suggests that task performance provides a unique contribution to activity changes that is not captured by age alone. Cortical thickness explained additional variance above age in activity in SMA only. The latter finding is consistent with earlier studies that showed that DLPFC and SPC structure could not explain age differences in neural activity (Haier et al., 2009; Squeglia et al., 2013). The relation between SMA cortical thickness and neural activity was positive, such that increased cortical thickness (i.e., less mature) is associated with increased activity. This is surprising given that cortical thickness decreased with age. Interestingly, however, our findings are consistent with a study in adult participants by Hegarty et al. (2012) who also found a positive relation between cortical thickness and activity in SMA in an inhibition paradigm, but not in other prefrontal areas. This relation should be addressed in more detail in future research.

4.3. Reliability of neural activity across child and adolescent development

Finally, in this study we measured consistency over time of activity in the frontoparietal network within individuals. There is a growing number of longitudinal studies in adults that examined test-retest reliability in the cognitive control network. The results indicate that over periods of two weeks, reliability is good in DLPFC and SPC during an *n*-back working memory task (Plichta et al., 2012), there is modest to good reliability of DLPFC activity in visual working memory over a period of three months (Zanto et al., 2014) and there is modest to good reliability of DLPFC and SPC in a feedback monitoring task over a period of three years (Koolschijn et al., 2011). The question of reliability in developmental populations has not been consistently examined, but a recent study suggested fair reliability in DLPFC, SPC and ACC in 123 participants between ages 9 to 29 in an oculomotor inhibition task (Ordaz et al., 2013). The current findings demonstrate that in a large sample of 208 participants, reliability across two years is fair to good across ages 8–27 years in bilateral DLPFC, SPC, SMA and ACC. The relatively high reliability of cortical regions is consistent with prediction studies that have shown that future academic achievement can be predicted by activity in SPC one year earlier (Emerson and Cantlon, 2014) and two years earlier (Dumontheil and Klingberg, 2012) during working memory tasks, suggesting that activity in DLPFC and SPC is related to future cognitive outcome.

4.4. Limitations

A limitation of this study is that the participants who were excluded for the cortical thickness analyses due to lower quality data ($n = 28$), were younger on average compared to the included participants, which is unfortunately a common problem in developmental fMRI studies. However, the remaining sample for cortical thickness was still large ($n = 180$; ages 8–27) and this was one of the first large-scale developmental longitudinal studies assessing the contribution of structural maturation to development of brain function. A second limitation of this study is that the feedback learning task was relatively easy and performance was near-perfect for older participants. Future studies should investigate whether the results from this feedback learning paradigm can be replicated using other and more cognitively taxing tasks measuring cognitive control. Third, older participants performed better on the learning task and needed fewer trials to complete the task, possibly confounding our findings. Future studies could focus on performance-controlled paradigms where tasks are manipulated to result in equal performance in children and adults. On the other hand, tasks resulting in age-related performance differences more closely resemble real-life situations. Finally, we only investigated whether time-related changes in neural activity covaried with the factors age, task performance, working memory and cortical thickness. Many other factors may contribute to individual differences in changes over time in activity, such as increased response inhibition, increased motivation or concentration, which should be investigated in future research.

4.5. Conclusions and future directions

This study moved beyond prior cross-sectional comparisons by fitting growth curves based on longitudinal data and thereby improved power for detecting developmental change. An interesting direction for future research is to test how neural activity predicts future changes in behavior. A prior study highlighted the important role of the parietal cortex for predicting future behavior such as academic performance (Dumontheil and Klingberg, 2012). However, other studies showed that activity in the frontoparietal network correlated with working memory across sessions, but activity in subcortical areas (basal ganglia and thalamus) predicted future working memory performance (Darki and Klingberg, 2014; Ullman et al., 2014).

In future studies it will be of interest to follow individuals in this study for a three or even more time points. With two time points, it was only possibly to investigate nonlinear patterns on the group level. In addition, an interesting future direction would be to examine whether and how baseline activity in the frontoparietal network predicts future behavioral outcomes. An important question for future research is to examine not only the spatial, but also the temporal dynamics of cognitive control, such as with event-related potentials. Prior studies have highlighted the feasibility of this approach in young children (Eppinger et al., 2009), and this will allow for the investigation of fast evaluative processes.

Taken together, this study accentuates the important role of the emerging frontoparietal network in child and adolescent development. With regard to dual-process models of adolescent development, this study provides evidence against a simple linear development of the frontoparietal network and highlights the need for further large-scale longitudinal studies to test adolescent development more reliably.

References

Braams, B.R., van Duijvenvoorde, A.C.K., Peper, J.S., Crone, E.A., 2015. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. J. Neurosci. 35, 7226–7238.

Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using the MarsBar toolbox for SPM 99. Neuroimage 16, S497.

Carter, C.S., van Veen, V., 2007. Anterior cingulate cortex and conflict detection: an update of theory and data. Cogn. Affect. Behav. Neurosci. 7, 367–379.

Case, R., 1992. The role of the frontal lobes in the regulation of cognitive development. Brain Cogn. 20, 51–73.

Church, J.A., Petersen, S.E., Schlaggar, B.L., 2010. The Task B problem and other considerations in developmental functional neuroimaging. Hum. Brain Mapp. 31, 852–862.
Cicchetti, D.V., 2001. The precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. J. Clin. Exp. Neuropsychol. 23, 695–700.
Cocosco, C.A., Kollokian, V., Kwan, R.K.S., Evans, A.C., 1997. Online interface to a 3D MRI simulated brain database. Neuroimage 5, S425.
Cohen, A.O., Breiner, K., Steinberg, L., Bonnie, R.J., Scott, E.S., Taylor-Thompson, K.A., . . . Casey, B.J., 2016. When is an adolescent an adult? assessing cognitive control in emotional and nonemotional contexts. Psychol. Sci., http://dx.doi.org/10.1177/0956797615627625.
Crone, E.A., Dahl, R.E., 2012. Understanding adolescence as a period of social–affective engagement and goal flexibility. Nat. Rev. Neurosci. 13, 636–650.
Crone, E.A., Zanolie, K., Van Leijenhorst, L., Westenberg, P.M., Rombouts, S.A., 2008. Neural mechanisms supporting flexible performance adjustment during development. Cogn. Affect. Behav. Neurosci. 8, 165–177.
Crone, E.A., 2014. The role of the medial frontal cortex in the development of cognitive and social-affective performance monitoring. Psychophysiology 51, 943–950.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9, 179–194.
Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–114.
Darki, F., Klingberg, T., 2014. The role of fronto-parietal and fronto-striatal networks in the development of working memory: a longitudinal study. Cereb. Cortex 25, bht352.
Dumontheil, I., Klingberg, T., 2012. Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. Cereb. Cortex 22, 1078–1085.
Dumontheil, I., Houlton, R., Christoff, K., Blakemore, S.-J., 2010. Development of relational reasoning during adolescence. Dev. Sci. 13, F15–24.
Emerson, R.W., Cantlon, J.F., 2014. Continuity and change in children's longitudinal neural responses to numbers. Dev. Sci. 18, 314–326.
Eppinger, B., Mock, B., Kray, J., 2009. Developmental differences in learning and error processing: evidence from ERPs. Psychophysiology 46, 1043–1053.
Ernst, M., Pine, D.S., Hardin, M., 2006. Triadic model of the neurobiology of motivated behavior in adolescence. Psychol. Med. 36, 299–312.
Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. U. S. A. 97, 11050–11055.
Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., . . . Kennedy, D., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.
Fjell, A.M., Walhovd, K.B., Westlye, L.T., Ostby, Y., Tamnes, C.K., Jernigan, T.L., . . . Dale, A.M., 2010. When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies. Neuroimage 50, 1376–1383.
Geier, C.F., Garver, K., Terwilliger, R., Luna, B., 2009. Development of working memory maintenance. J. Neurophysiol. 101, 84–99.
Haier, R.J., Karama, S., Leyba, L., Jung, R.E., 2009. MRI assessment of cortical thickness and functional activity changes in adolescent girls following three months of practice on a visual-spatial task. BMC Res. Notes 2, 174.
Harms, M.P., Wang, L., Csernansky, J.G., Barch, D.M., 2013. Structure-function relationship of working memory activity with hippocampal and prefrontal cortex volumes. Brain Struct. Funct. 218, 173–186.
Hegarty, C.E., Foland-Ross, L.C., Narr, K.L., Townsend, J.D., Bookheimer, S.Y., Thompson, P.M., Altshuler, L.L., 2012. Anterior cingulate activation relates to local cortical thickness. Neuroreport 23, 420–424.
Huizinga, M., Dolan, C.V., van der Molen, M.W., 2006. Age-related change in executive function: developmental trends and a latent variable analysis. Neuropsychologia 44, 2017–2036.
Johnson, M.H., 2011. Interactive specialization: a domain-general framework for human functional brain development? Dev. Cogn. Neurosci. 1, 7–21.
Kleibeuker, S.W., De Dreu, C.K.W., Crone, E.A., 2013. The development of creative cognition across adolescence: distinct trajectories for insight and divergent thinking. Dev. Sci. 16, 2–12.
Klingberg, T., Forssberg, H., Westerberg, H., 2002. Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. J. Cogn. Neurosci. 14, 1–10.
Koolschijn, P.C.M.P., Crone, E.A., 2013. Sex differences and structural brain maturation from childhood to early adulthood. Dev. Cogn. Neurosci. 5, 106–118.
Koolschijn, P.C.M.P., Schel, M.A., de Rooij, M., Rombouts, S.A., Crone, E.A., 2011. A three-year longitudinal functional magnetic resonance imaging study of performance monitoring and test-retest reliability from childhood to early adulthood. J. Neurosci. 31, 4204–4212.
Koolschijn, P.C.M.P., Peper, J.S., Crone, E.A., 2014. The influence of sex steroids on structural brain maturation in adolescence. PLoS One.
Larson, G.E., Merritt, C.R., Williams, S.E., 1988. Information processing and intelligence: some implications of task complexity. Intelligence 12, 131–147.
Lu, L.H., Dapretto, M., O'Hare, E.D., Kan, E., McCourt, S.T., Thompson, P.M., . . . Sowell, E.R., 2009. Relationships between brain activation and brain structure in normally developing children. Cereb. Cortex 19, 2595–2604.
Mars, R.B., Coles, M.G., Grol, M.J., Holroyd, C.B., Nieuwenhuis, S., Hulstijn, W., Toni, I., 2005. Neural dynamics of error processing in medial frontal cortex. Neuroimage 28, 1007–1013.
Ordaz, S.J., Foran, W., Velanova, K., Luna, B., 2013. Longitudinal growth curves of brain function underlying inhibitory control through adolescence. J. Neurosci. 33, 18109–18124.
Paulsen, D.J., Hallquist, M.N., Geier, C.F., Luna, B., 2015. Effects of incentives, age: and behavior on brain activation during inhibitory control: a longitudinal fMRI study. Dev. Cogn. Neurosci. 11, 105–115.
Peters, S., Braams, B.R., Raijmakers, M.E.J., Koolschijn, P.C.M.P., Crone, E.A., 2014a. The neural coding of feedback learning across child and adolescent development. J. Cogn. Neurosci. 26, 1705–1720.
Peters, S., Koolschijn, P.C.M.P., Crone, E.A., Van Duijvenvoorde, A.C.K., Raijmakers, M.E.J., 2014b. Strategies influence neural activity for feedback learning across child and adolescent development. Neuropsychologia 62, 365–374.
Pfeifer, J.H., Allen, N.B., 2012. Arrested development? Reconsidering dual-systems models of brain function in adolescence and disorders. Trends Cogn. Sci. 16, 322–329.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2007. NLME: Linear and nonlinear mixed effects models. R package version 3. 1–122. http://CRAN.R-project.org/package=nlme.
Plichta, M.M., Schwarz, A.J., Grimm, O., Morgen, K., Mier, D., Haddad, L., . . . Meyer-Lindenberg, A., 2012. Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. Neuroimage 60, 1746–1758.
Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–1418.
Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., Nieuwenhuis, S., 2004. The role of the medial frontal cortex in cognitive control. Science 306, 443–447.
Rubia, K., Smith, A.B., Woolley, J., Nosarti, C., Heyman, I., Taylor, E., Brammer, M., 2006. Progressive increase of frontostriatal brain activation from childhood to adulthood during event-related tasks of cognitive control. Hum. Brain Mapp. 27, 973–993.
Ségonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. Neuroimage 22, 1060–1075.
Shaw, P., Malek, M., Watson, B., Greenstein, D., de Rossi, P., Sharp, W., 2013. Trajectories of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. Biol. Psychiatry 74, 599–606.
Somerville, L.H., Casey, B.J., 2010. Developmental neurobiology of cognitive control and motivational systems. Curr. Opin. Neurobiol. 20, 236–241.
Somerville, L.H., Jones, R.M., Ruberry, E.J., Dyke, J.P., Glover, G., Casey, B.J., 2013. The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. Psychol. Sci. 24, 1554–1562.
Squeglia, L.M., McKenna, B.S., Jacobus, J., Castro, N., Sorg, S.F., Tapert, S.F., 2013. BOLD response to working memory not related to cortical thickness during early adolescence. Brain Res. 1537, 59–68.
Steinberg, L., 2008. A social neuroscience perspective on adolescent risk-taking. Dev. Rev. 28, 78–106.
Talairach, J., Tourneaux, P., 1988. Co-planar Stereotactic Atlas of the Human Brain. Thieme, New York.
Tamnes, C.K., Østby, Y., Fjell, A.M., Westlye, L.T., Due-Tønnessen, P., Walhovd, K.B., 2010. Brain maturation in adolescence and young adulthood: regional age-related changes in cortical thickness and white matter volume and microstructure. Cereb. Cortex 20, 534–548.
Teslovich, T., Mulder, M., Franklin, N.T., Ruberry, E.J., Millner, A., Somerville, L.H., . . . Casey, B.J., 2014. Adolescents let sufficient evidence accumulate before making a decision when large incentives are at stake. Dev. Sci. 17, 59–70.
Thomason, M.E., Race, E., Burrows, B., Whitfield-Gabrieli, S., Glover, G.H., Gabrieli, J.D.E., 2009. Development of spatial and verbal working memory capacity in the human brain. J. Cogn. Neurosci. 21, 316–332.
Ullman, H., Almeida, R., Klingberg, T., 2014. Structural maturation and brain activity predict future working memory capacity during childhood development. J. Neurosci. 34, 1592–1598.
Van Duijvenvoorde, A.C.K., Zanolie, K., Rombouts, S.A., Raijmakers, M.E., Crone, E.A., 2008. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. J. Neurosci. 28, 9495–9503.
Van den Bos, Güroğlu, B., van den Bulk, B.G., Rombouts, S.A.R.B., Crone, E.A., 2009. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. Front. Hum. Neurosci. 3, 52.
Velanova, K., Wheeler, M.E., Luna, B., 2008. Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. Cereb. Cortex 18, 2505–2522.
Wendelken, C., O'Hare, E.D., Whitaker, K.J., Ferrer, E., Bunge, S.A., 2011. Increased functional selectivity over development in rostrolateral prefrontal cortex. J. Neurosci. 31, 17260–17268.
Zanolie, K., Van Leijenhorst, L., Rombouts, S.A., Crone, E.A., 2008. Separable neural mechanisms contribute to feedback processing in a rule-learning task. Neuropsychologia 46, 117–126.
Zanto, T.P., Pa, J., Gazzaley, A., 2014. Reliability measures of functional magnetic resonance imaging in a longitudinal evaluation of mild cognitive impairment. Neuroimage 84, 443–452.