Ongoing neural development of affective theory of mind in adolescence

Nora C. Vetter,[1] **Sarah Weigelt,**[2,3] **Katrin Döhnel,**[4] **Michael N. Smolka,**[1] **and Matthias Kliegel**[5]

[1]Department of Psychiatry and Psychotherapy, and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany, [2]Massachusetts Institute of Technology, Department of Brain and Cognitive Science and McGovern Institute for Brain Research, Cambridge, MA, USA, [3]Department of Psychology, Ruhr-Universität Bochum, Bochum, Germany, [4]Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany, and [5]Department of Psychology, University of Geneva, Geneva, Switzerland

Affective Theory of Mind (ToM), an important aspect of ToM, involves the understanding of affective mental states. This ability is critical in the developmental phase of adolescence, which is often related with socio-emotional problems. Using a developmentally sensitive behavioral task in combination with functional magnetic resonance imaging, the present study investigated the neural development of affective ToM throughout adolescence. Eighteen adolescent (ages 12–14 years) and 18 young adult women (aged 19–25 years) were scanned while evaluating complex affective mental states depicted by actors in video clips. The ventromedial prefrontal cortex (vmPFC) showed significantly stronger activation in adolescents in comparison to adults in the affective ToM condition. Current results indicate that the vmPFC might be involved in the development of affective ToM processing in adolescence.

Theory of Mind (ToM), the ability to infer others' mental states (Perner, 1991; Frith and Frith, 2003) can be divided into: (i) cognitive ToM encompassing inferences about 'cold' mental states such as intentions and beliefs and (ii) affective ToM encompassing inferences about 'hot' mental states, that is, emotions (Shamay-Tsoory *et al.*, 2010). Well-functioning skills of both ToM aspects are much needed in the developmental period of adolescence (Steinberg and Morris, 2001). Surprisingly, studies have only recently begun to shed light on the development of cognitive and affective ToM across adolescence (Blakemore, 2008). Ongoing refinement of cognitive and affective ToM across adolescence was indicated by first studies, mostly in reaction time (RT) measures (Choudhury *et al.*, 2006; Keulers *et al.*, 2010). An improvement in accuracy was demonstrated on both cognitive and affective ToM within the same sample of adolescents (Vetter *et al.*, 2012). Moreover, this study suggested greater age differences between adolescents and adults for affective ToM compared with cognitive ToM. This finding together with previous developmental studies indicates an extended developmental trajectory of affective ToM in contrast to cognitive ToM (e.g. Baron-Cohen *et al.*, 1999). Conceptually, affective ToM seems to be particularly complex requiring the integration of both cognitive ToM and empathy. Therefore, and while previous neural studies have largely focused on the development of cognitive ToM (for a review, see Blakemore, 2008), the current study focused on the development of affective ToM in adolescence. The 'Faces test' (Golan *et al.*, 2006), a developmentally sensitive paradigm (Vetter *et al.*, 2012, 2013) presenting complex emotional mental states was employed. By using this ecological valid measure of affective ToM

development in adolescence, the current functional magnetic resonance imaging (fMRI) study investigated the neural basis of affective ToM development across adolescence.

The adult neural ToM network (Van Overwalle, 2009) has consistently been shown to comprise the posterior superior temporal sulcus (pSTS; Puce *et al.*, 1998), the temporal pole (TP; Frith and Frith, 2003) and the temporo-parietal junction (TPJ; Saxe and Kanwisher, 2003). These regions have also been confirmed for affective ToM (Hynes *et al.*, 2006; Völlm *et al.*, 2006; Sebastian *et al.*, 2012).

Another important ToM region is the medial prefrontal cortex (mPFC; Van Overwalle, 2009; Abu-Akel and Shamay-Tsoory, 2011). With respect to affective ToM, especially the ventromedial PFC (vmPFC) has been observed. Strongest evidence comes from findings of vmPFC-lesioned patients showing deficits specifically for affective ToM. Concurrently, these patients appear to be impaired on recognizing affective mental states such as emotions (Heberlein *et al.*, 2008), a faux pas or irony (Stone *et al.*, 1998; Shamay-Tsoory *et al.*, 2006; Shamay-Tsoory and Aharon-Peretz, 2007). Corroborating these findings anatomically, the vmPFC has strong connections with affect-processing regions such as the amygdala (Bandler *et al.*, 2000; Price, 2007). However, functional neuroimaging studies appear to support the importance of vmPFC for affective ToM only partly. Whereas Hynes *et al.* (2006) found differential vmPFC activity for affective ToM, other authors observed activity in the dorsomedial PFC (dmPFC; Völlm *et al.*, 2006) or a cluster reaching from dorso- to ventromedial PFC (Sebastian *et al.*, 2012).

Regarding developmental findings on affective ToM processing, results of brain regions showing a stronger activation in adolescents in comparison to adults are 3-fold: while one study observed dmPFC involvement (Wang *et al.*, 2006), another found both the dmPFC and the vmPFC (Gunther Moor *et al.*, 2012) and a third one observed an activation of the vmPFC (Sebastian *et al.*, 2012). Moreover, additional regions were found such as the right pSTS (Wang *et al.*, 2006) or the right TP (Gunther Moor *et al.*, 2012). Thus, until now, there is no clear-cut picture as to which neuronal structures underlie the continued development of affective ToM.

Heterogeneous findings of the aforementioned developmental studies might be due to three possible reasons. First, studies investigated

different aspects of affective ToM, requiring more or less putting one-self into other's emotional shoes (i.e. empathy). Wang *et al.* (2006) asked whether statements were meant sincere or ironic in cartoons, which might require less empathy and instead rather perspective taking, that is, cognitive ToM. In contrast, in vignettes, participants had to choose the correct reaction of one character to her companion's affective state (Sebastian *et al.*, 2012), which might require more empathy. In eye regions, participants needed to evaluate the correct affective expression probably also requiring more empathic processes (Gunther Moor *et al.*, 2012).

Second, studies rather employed children's tasks, which may not be performance sensitive for adolescents. For example, Wang *et al.* used a children's task and found ceiling effects in accuracy. Supporting this notion, significant behavioral differences were only observed in Sebastian *et al.* using more complex social material: adolescents performed lower than adults in the vignette paradigm.

Third, another reason for the heterogeneous results for neural affective ToM development seems to be large differences between investigated age spans: while Wang *et al.* (2006) included 9- to 14-year olds, Sebastian *et al.* (2012) investigated 11- to 16-year-old adolescents. Thus, these studies recruited adolescent groups with a wide age range of 5 years. However, it is desirable to trace developmental changes in narrow age ranges given the gross developmental changes in brain structure observed during adolescence (Giedd, 2008). This was done by Gunther Moor *et al.* (2012) who differentiated between early (10–12 years) and middle adolescents (14–16 years) in narrow age clusters.

The current study aimed at extending previous findings by addressing these challenges. First, the affective ToM paradigm required more robust empathy skills because it assesses the ability to evaluate subtle mental states in realistic video clips. Second, this paradigm is developmentally sensitive since performance differences in adolescents and adults have been shown (Vetter *et al.*, 2013). Moreover, the method of performance matching was applied. Controlling for performance systematically has become a demand of neurodevelopmental studies (Schlaggar *et al.*, 2002; Church *et al.*, 2010). Otherwise, it is unclear whether neural differences are due to age or just due to performance differences (Ernst and Mueller, 2008). Adolescent participants were matched to adults with similar performance leading to comparable performance across age groups (e.g. Schlaggar *et al.*, 2002). Although in other developmental areas, performance matching has been employed successfully (Schlaggar *et al.*, 2002; Braet *et al.*, 2009), to our knowledge, the current study is the first in the area of developing ToM employing a performance-matching strategy. Third, narrow age ranges for both the adolescent (12–14 years) and the adult group (19–25 years) were chosen. By using a developmentally sensitive affective ToM paradigm requiring empathy in narrow age groups and by using a performance-matching procedure, the aim of the current fMRI study was to further explore age-related changes in functional activity associated with affective ToM processing in adolescence relative to adulthood. We hypothesized to find a stronger vmPFC activation in adolescents in comparison to adults because the vmPFC might be most sensitive for affective ToM development as indicated by lesion studies and first developmental studies.

METHODS
Participants
Originally, 32 adolescent and 20 adult female volunteers were recruited via flyers (preuniversity education and undergraduate university students). We measured only females since structural and functional brain development is related to gender (Giedd, 2008). Adolescents received monetary compensation and university students participated for course credit. Informed consent was obtained from each participant

and additionally for adolescents from one of their legal guardians. The study was approved by the local ethics committee. Three adolescents and one adult were excluded due to excessive movement and one adolescent and one adult due to technical problems. This resulted in 28 adolescent and 18 adult participants with no record of neurological or psychiatric illness. All participants were right handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), spoke German as their first language and had normal or corrected to normal vision. Performance in terms of accuracy in the affective ToM condition differed significantly across groups, $t(44) = -2.76$, $P < 0.01$; $mean_{adolescents} = 79.83$, $s.d._{adolescents} = 10.58$; $mean_{adults} = 87.27$, $s.d._{adults} = 5.24$. In order to achieve equal performance on the affective ToM task in both groups, adolescents with the highest performance in the affective ToM task were chosen to match performance of the adult age group (Table 1). The performance-matched groups contained 18 adolescents (range 12.07–14.61 years) and 18 adults (range 19.1–25.77 years; Table 2). According to the Pubertal Development Scale (Petersen *et al.*, 1988) used in a German version (Watzlawik, 2009), 22.2% ($n = 4$) of the adolescent sample was midpubertal and 77.7% ($n = 14$) late pubertal, which is in line with findings by Watzlawik. Given the small group of midpubertal adolescents, we did not investigate neural changes due to pubertal status, which could be aimed at in future studies. Groups did not differ with respect to socioeconomic status and age-corrected verbal and non-verbal abilities (Table 2).

Stimuli, design and procedure
We developed an affective ToM task adapted from the 'facial scale' of the Cambridge Mindreading Face-Voice Battery (Golan *et al.*, 2006) and added a physical control task. The facial scale has been employed behaviorally with adolescents of the target age group to ensure that it covered the dynamic range of performances in the adolescent group (Vetter *et al.*, 2013). Silent film clips of different actors expressing mental states in the face and torso (from the shoulders upward) were presented (Figure 1). In the affective ToM task, participants were instructed to choose the adjective that best describes the actor's mental state out of four affective adjectives. Different target and distractor adjectives were used for each film clip. Examples of adjectives are resentful, uneasy and subdued. In the physical control task, participants were instructed to report on either the color of the actor's T-shirt, hair or skin. Each correctly solved film clip yielded one point, resulting in a maximum raw score of 48 for the affective ToM; respectively, the physical control condition. Performance at chance would be a raw score of 12.

Forty-eight film clips were shown once for the affective ToM and the same 48 film clips were shown once for the physical control task. In the physical control condition, each question (color of T-shirt, hair or skin) was given 16 times. The film clips were controlled systematically in terms of gender and age group of the actor; there were three age groups (adolescents, young adults, middle- to old-aged adults) so that 16 actors (8 females and 8 males) per stimulus age group were depicted. The film clips were presented in a pseudorandom order to assure that an individual film clip was not immediately repeated.

Each trial (Figure 1) began with the instruction screen displayed for 1.5 s consisting of a cue word ('emotion' for the affective ToM condition, respectively, 'body' for the physical control condition). The cue word 'body' signaled to concentrate on the three physical features of the person leaving open which physical component would be demanded until the choice screen, that specified the question, was presented. This was to assure that the participant concentrated on the film clip continuously since she had to consider three different features. After the instruction screen, exponentially jittered interstimulus intervals (ISIs) were employed varying randomly from 2 to 6 s (Serences, 2004). This enabled the separation of the neural response of the

Table 1 Means (s.d.'s) for percentage of correct responses and RTs

	Mean (s.d.)	
	Adults, $N = 18$ females	Adolescents, $N = 18$ females
Percentage correct		
Affective ToM	87.27 (5.24)	85.53 (5.55)
Physical control	88.08 (6.5)	84.37 (3.93)
RT		
Affective ToM	2517 (301)	2649 (398)
Physical control	1810 (190)	1971 (201)

RT is given in milliseconds for correct-only trials. Groups did not differ on percentage of correct responses or RT.

Table 2 Means (s.d.'s) for sample characteristics

	Adults $N = 18$ females	Adolescents $N = 18$ females	Age group comparison	
	Mean (s.d.)	Mean (s.d.)	t	P
Age	21.24 (1.55)	13.7 (0.77)		
Verbal ability	14.67 (2.22)	13.89 (2.11)	−1.08	0.29
Non-verbal ability	11.94 (1.62)	12.11 (2.03)	0.27	0.79
Socioeconomic status	15.39 (3.78)	14.08 (5.11)	−0.87	0.39

Verbal and non-verbal ability were measured with the subtests of the Wechsler Adult Intelligence Scale for adult participants (German adaptation; von Aster et al., 2007) and the Wechsler Intelligence Scale For Children for adolescent participants (WISC-IV, German adaptation; Petermann and Petermann, 2007). For both verbal and non-verbal ability, scores are age corrected (mean = 10, s.d. = 3). Calculation of socioeconomic status included parents' school education, professional education, recent professional status and family income following the procedure suggested by Winkler and Stolzenberg (2009). Scores for mothers and fathers were averaged into a family-based measure of socioeconomic background. The score ranges from 3 to 21 with higher values indicating higher socioeconomic status.

instruction from that of the film clip. The ISI was followed by the film clip lasting 5.5 s and by the choice screen presented for 6.5 s. During the presentation of the choice screen, the participant gave an answer via button press. After the button press, participant's choice was highlighted with a color change of the selected word, which remained colored until the end of the 6.5 s period. There was no feedback given to the participant. Each trial varied from 15.5 to 19.5 s and the whole functional run lasted about 30 min. Behavioral data were collected by ResponseGrips (©NordicNeuroLab) with two buttons for each hand. The correct response alternative was equally distributed among the buttons. Task presentation and recordings of the behavioral responses were performed using Presentation® software (version 11.1, Neurobehavioral Systems, Inc., Albany, CA, USA).

The scanning session was preceded by a short introductory session outside the scanner, followed by a practice session inside the scanner (including eight additional film clips not included in the main task). In addition, adolescent participants were made familiar with the scanning environment by use of a mock scanner (Galvan et al., 2012). A hand-out was provided before the scan containing all affective adjectives to ensure that they were known to the participants. Participants were instructed to report if they did not know any of them. The experimenter then gave standardized definitions.

Moreover, partial trials (Ollinger et al., 2001) were employed for separately estimating the hemodynamic response to neural events occurring in a fixed sequence (i.e. film clip followed by the choice screen). Therefore, six additional film clips followed by a fixation cross, presented twice per condition were employed. However, analysis concentrated on the film clip.

Statistical analysis of behavioral data

Statistical analyses were performed using SPSS for Windows (Version 18) applying mixed model repeated measures ANOVAs with a 2×2

factorial design: age group (adolescents, adults) as the between-subjects factor and condition (affective ToM, physical control) as the within-subjects factor. The percentage of correct responses and the RTs were used as the dependent variables and a threshold of $P < 0.05$ was applied.

Functional imaging

Image acquisition

Scanning was performed with a 3 T whole-body MR tomograph (Magnetom TRIO, Siemens, Erlangen, Germany) equipped with a 12-channel head coil. For functional imaging, a standard echo planar imaging sequence was used (TR: 2410 ms; TE: 25 ms; flip angle: 80°). fMRI scans were obtained from 42 transversal slices, tilted up 30° clockwise from the anterior commissure–posterior commissure line to improve signal in the orbitofrontal cortex and minimize susceptibility artifacts. A thickness of 2 mm (1 mm gap), an field of view (FOV) of 192×192 mm and an in-plane resolution of 64×64 pixels resulted in a voxel size of $3 \times 3 \times 3$ mm. Only marginal sections of the most superior part of the parietal cortex and the most inferior part of the cerebellum were omitted for subjects with a larger brain that did not fit into the field of view. All ToM relevant regions such as the TPJ were included. Moreover, a 3D T1-weighted magnetization prepared rapid acquisition gradient echo (MPRAGE) image data set was acquired (TR: 1900 ms; TE: 2.26 ms; field of view (FOV): 256×256 mm; 176 slices; $1 \times 1 \times 1$ mm voxel size; flip angle = 9°). Images were presented via magnet-compatible goggles (VisuaStim™, Resonance Technology, CA, USA or Nordic Neurolab, Bergen, Norway).

fMRI data analysis

Functional images were preprocessed and statistically analyzed using SPM 8 (Wellcome Department of Imaging Neuroscience, London, UK). For each participant, functional images were first slice-time corrected by using the middle slice as reference, then realigned to the mean image by 6° rigid spatial transformation (Friston et al., 1995), spatially normalized (Ashburner and Friston, 1999) to the standard space defined by the Montreal Neurological Institute (MNI) template and smoothed with a Gaussian kernel of 8 mm at full-width half maximum. Adolescents and adults did not differ regarding movement parameters.

In the first-level analysis, a fixed effects analysis was computed for each subject on the basis of the general linear model (GLM) within each voxel of the whole brain. The analysis focused on amplitude changes in the hemodynamic response function associated with affective ToM processing in the experimental film clip condition contrasted with processing physical appearance in the control film clip condition (Figure 1). We did not examine activation during choice because it might be confounded by motor activation and reading/mere choice processes. The GLM included as the main regressor of interest the film clip in the two conditions modeled with its duration of 5500 ms. In addition, the instruction period was modeled with 1500 ms as a regressor of no interest. Furthermore, the response phase was split into three separate regressors of no interest. This enabled to estimate the underlying psychological processes more accurately since they were assumed to differ. These regressors comprised the choice screen (duration = RT), the button press (event with no duration) and the color change of choice feedback (duration = 6500 ms minus RT). All regressors were modeled as boxcar functions convolved with a canonical hemodynamic response function (except the button press modeled as a stick function). Additionally, six subject-specific movement regressors were included as covariates of no interest. Each component of the model served as a regressor in a multiple regression analysis. A high-pass filter with cut-off 128 s for removing low-frequency physiological
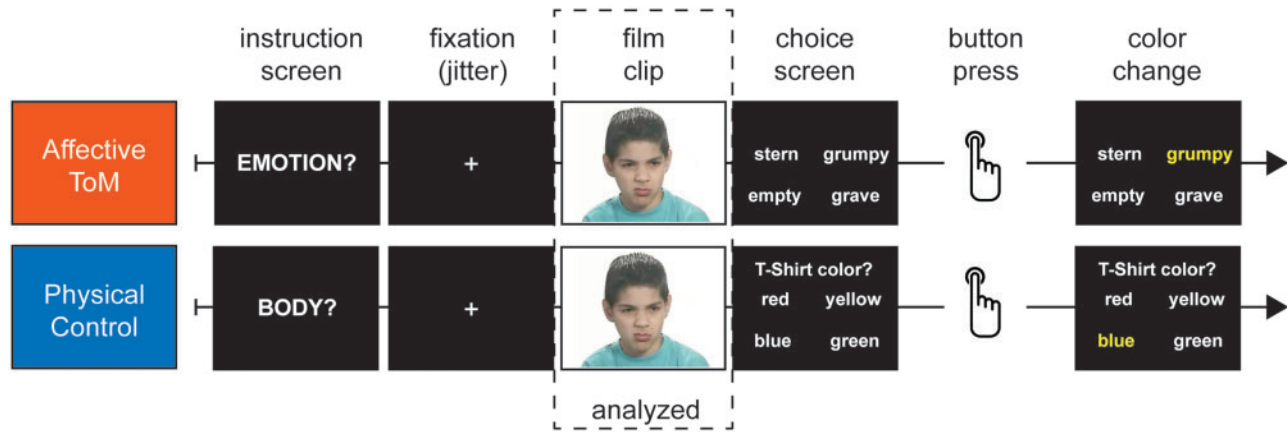
Fig. 1 Example trials for the affective ToM and the physical control condition.

noise and an AR(1) model for the residual temporal autocorrelation were employed (Henson, 2006). Statistical parametric maps (SPMs) were generated for each subject by *t*-statistics derived from contrasts utilizing the HRF. Three contrasts of interest were computed within each subject: affective ToM minus baseline (Contrast 1), physical control minus baseline (Contrast 2) and affective ToM minus physical control (Contrast 3). The first-level contrast images from the weighted beta images were introduced into second-level whole-brain random-effects analysis to allow for population inference.

In order to investigate the main effect of condition, an ANOVA was computed using a 2 × 2 flexible factorial model[1] with the factors group (adolescents, adults) and condition (using Contrasts 1 and 2). A subject factor was used in the flexible factorial model. The resulting set of significant voxel values constituted an SPM map. The SPM maps were thresholded at $P \leq 0.001$ (uncorrected voxel level). We report regions that survive a threshold of $P \leq 0.05$ (corrected for multiple tests on the cluster level). All brain coordinates are reported in MNI atlas space.

We followed three streams of analysis: first, related to our hypothesis, we analyzed the vmPFC and dmPFC, which have previously most robustly shown developmental effects in affective ToM studies. We expected to find clusters of vmPFC and dmPFC in the main effect of condition. We analyzed these clusters for a significant group (adolescents, adults) × condition (affective ToM, physical control) interaction. Second, we analyzed the regions of interest (ROIs) of TP and pSTS, which have shown age effects for affective ToM in Wang *et al.* (2006) and Gunther Moor *et al.* (2012) for interaction of group by condition. Third, we computed a whole-brain SPM and analyzed the interaction of group by condition.

In detail, we explored the response profile of the resulting clusters of vmPFC and dmPFC from the main effect of condition (affective ToM > physical control). Therefore, masks of these clusters were created. Applying these masks percent-signal change was extracted with rfxplot (Glaescher, 2009). We calculated an ANOVA (condition × group) of the percent-signal change values in the vmPFC and dmPFC cluster.

We further converted the pSTS peak voxel of Wang *et al.* from Talairach (*x*, *y*, *z* = 42, −44, 16) to MNI space (*x*, *y*, *z* = 47, −45, 17) using the Tal2ICBM function (Lancaster *et al.*, 2007). Then, we used the peak voxels of the pSTS and the TP (*x*, *y*, *z* = 33, 12, −30; Gunther Moor *et al.*) and built a 10 mm sphere around it to analyze interaction of condition × group in the main effect of condition.

RESULTS

Behavioral results

Behavioral results are displayed in Table 1. Regarding response accuracy, the ANOVA revealed no main effects of condition [$F(1,34) = 0.029$, $P = 0.87$] or group [$F(1,34) = 3.37$, $P = 0.08$] and no significant group by condition interaction [$F(1,34) = 0.94$, $P = 0.34$]. For RTs, the ANOVA showed no main effect of group [$F(1,34) = 3.17$, $P = 0.08$] and no significant group by condition interaction [$F(1,34) = 0.09$, $P = 0.77$]. However, there was a significant main effect of condition [$F(1,34) = 211.61$, $P < 0.001$]. *Post hoc t*-tests revealed that this was driven by slower RTs in the affective ToM condition compared with the physical control condition [$t(35) = 14.74$, $P < 0.001$] across both groups.

fMRI results

Main effect of condition

There was no main effect of group in the full factorial model. The main effect of condition for affective ToM > physical control revealed activity across both groups in bilateral TPJ/pSTS, extending from middle to anterior STS and to the TP (Table 3). Furthermore, the inferior frontal gyrus, the ventral striatum, the superior frontal gyrus, the parahippocampal gyrus extending to the amygdala, the cuneus and the cerebellum were activated bilaterally. Additionally, the left thalamus was activated. Importantly, both the vmPFC and dmPFC were activated.

Percent-signal change of clusters vmPFC and dmPFC from the main effect of condition. Percent-signal change analysis of the vmPFC mask obtained from the main effect of condition revealed a significant interaction of condition by group, $F(1,34) = 9.83$, $P < 0.01$, Figure 2. *Post hoc t*-tests revealed that the interaction was driven by the adolescent group, showing significantly more activation during the affective ToM relative to the physical control condition, $t(17) = 6.14$, $P < 0.001$, while adults' activation in this area did also differ between conditions but less strongly, $t(17) = 2.53$, $P < 0.05$, Figure 2. Furthermore, a significant difference of affective ToM between adults and adolescents emerged: adolescents showed more activation in the affective ToM condition than adults, $t(34) = 3.51$, $P < 0.01$, Figure 2. There was no difference between adults' and adolescents' vmPFC activation for physical control, $t(34) = 0.73$, $P = 0.47$. The percent-signal change analysis in the dmPFC cluster showed that the interaction of condition by group was not significant, $F(1,34) = 0.003$, $P = 0.95$, Figure 2.

Table 3 Functional activity associated with the main effect of condition (affective ToM vs physical control)

Brain region	L/R	BA	Peak voxel (mm)			t-value	Cluster corrected P-value	Cluster size
			x	y	z			
vmPFC	L	11	0	40	−18	6.39	0.013	183
dmPFC	L	9	−8	58	32	5.92	0.007	218
Inferior frontal gyrus	L	47	−48	26	−4	15.74	<0.001	14 606
	R	45	58	26	12	8.71		Part of same cluster
TP	L	38	−48	14	−26	11.81		Part of same cluster
	R	38	48	14	−32	10.73		Part of same cluster
Anterior STS	L	21	−58	0	−16	10.84		Part of same cluster
	R	21	54	4	−22	10.32		Part of same cluster
Middle STS	R	22	48	−36	2	10.46		Part of same cluster
	L	22	−56	−40	4	8.16		Part of same cluster
TPJ/pSTS	L	21	−60	−48	8	10.56		Part of same cluster
	R	22	64	−50	12	6.76		Part of same cluster
Parahippocampal gyrus	L	28	−22	−14	−14	7.01		Part of same cluster
	R	28	20	−16	−16	6.53		Part of same cluster
Ventral striatum	L		−10	6	9	7.59		Part of same cluster
	R		14	10	0	6.64		Part of same cluster
Thalamus	L		−2	−6	8	8.55		Part of same cluster
Fusiform gyrus	L	36	−42	−38	−20	5.79	0.045	110
Cuneus/lingual gyrus	R	17	20	−92	−4	7.79	<0.001	591
Cerebellum	R		26	−86	−36	10.94	<0.001	538
	L		−20	−76	−38	11.72	<0.001	1779
Superior frontal gyrus	R	6	8	14	72	6.50	<0.001	656

Brodmann areas (BAs) are approximate. Some clusters showed activation in multiple brain regions and BAs. P < 0.05, corrected cluster level.
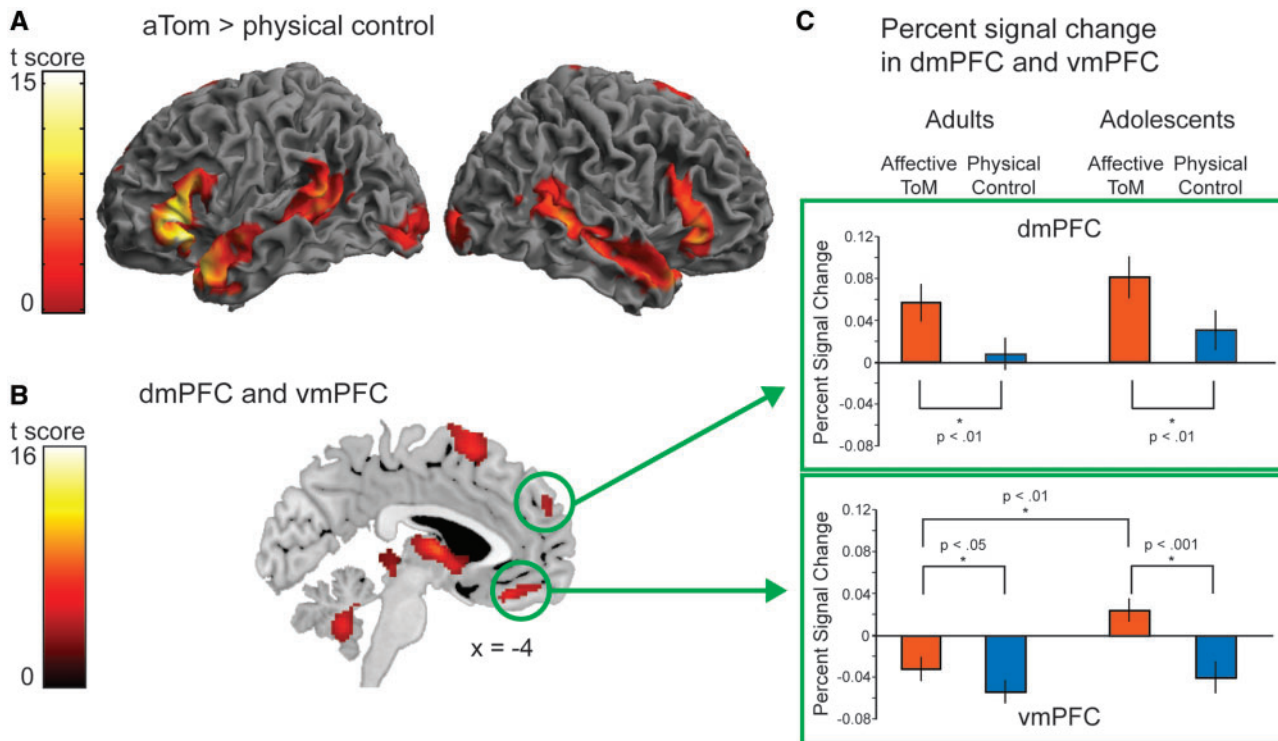

Fig. 2 (**A**) Left and right lateral renderings as well as a sagittal slice (**B**), overlaid on the MNI T1 template of the main effect of condition (affective ToM > physical control) in all 36 participants, thresholded at p < 0.001 uncorrected voxel level, P < 0.05 corrected cluster level and (**C**) analysis of percent-signal change of the vmPFC and dmPFC cluster taken from (B).

To directly test the differential involvement of the vmPFC and dmPFC in the development of affective ToM, we conducted a three-factorial ANOVA on the percent-signal change values with factors region (vmPFC, dmPFC), group (adolescents, adults) and condition (affective ToM, physical control). This revealed a significant main effect of region, $F(1,34) = 47.33$, $P < 0.001$, arising from the higher activation of dmPFC across groups and conditions. Importantly, it also revealed a significant region by condition by age group interaction, $F(1,34) = 5.28$, $P < 0.05$, arising from the difference of regions between groups only in the affective ToM condition.
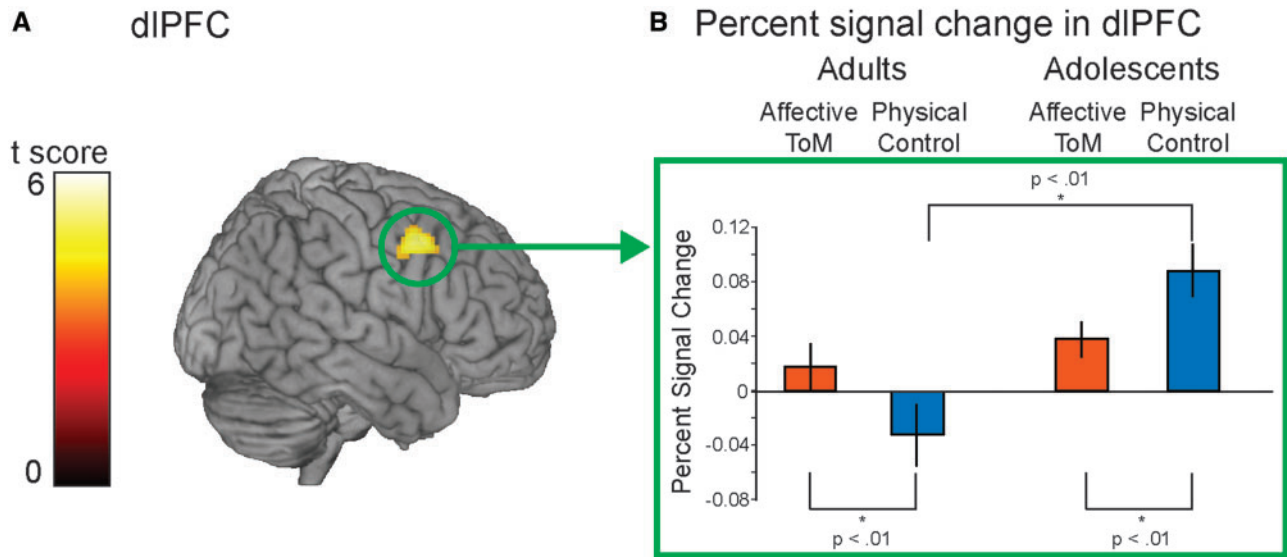
Fig. 3 (**A**) Result of the interaction group (adults *vs* adolescents) × condition (physical control *vs* affective ToM) in the dlPFC. (**B**) The shape of the interaction using the mask of the resulting cluster was further explored extracting percent-signal change.

Analyses of TP and pSTS ROIs. In the ROI analyses of the clusters derived from Gunther Moor *et al.* (2012) and Wang *et al.* (2006), there was no significant interaction of group × condition.

Whole-brain interaction of condition × group. For the interaction of group × condition in the direction of (adolescents > adults) × (affective ToM > physical control), there were no significant activations on a corrected cluster level of $P < 0.05$.

For the reverse interaction, adults showed higher activation in affective ToM *vs* physical control compared with adolescents in one cluster in the right dorsolateral PFC (dlPFC, *x, y, z* = 48, 8, 44; *k* = 237, *t* = 5.26, $P < 0.05$ corrected cluster level; Figure 3).

Post hoc t-tests conducted on the percent-signal change values showed that the interaction was driven by the difference of affective ToM and physical control in both age groups: adults showed a greater response of the dlPFC during affective ToM in contrast to physical control, *t*(17) = 3.3, $P < 0.01$, whereas adolescents did show the reverse, namely a lower response during affective ToM than physical control, *t*(17) = −3.41, $P < 0.01$ (Figure 3).

Overall, results did not differ when only those film clips that were rated correctly were included in the analyses.

Comparison of performance matched and non-matched groups. We additionally compared the matched and non-matched groups on behavioral measures (Supplementary Table S1). Taken together, groups did not differ, that is, the matched group does not seem to have better cognitive skills in general. By definition, groups differed in affective ToM performance. Thus, we analyzed fMRI data of the full sample of adolescents (Supplementary Tables S2 and S3). For the main effect of the condition, the overall pattern of activation (affective ToM > physical control; Supplementary Table S2) replicated findings of the performance-matched groups (Table 2). For the vmPFC, the full-sample analysis trended into the same direction, it just did not reach significance on a cluster corrected level ($P = 0.06$). Percent-signal change analysis of the ROI cluster resulting from the main effect in the matched sample revealed no significant interaction. Taken together, the vmPFC results were at trend (main effect), but did not fully replicate findings of the matched group (ROI analysis). In contrast, the

interaction of condition and group in the dlPFC held for the full sample (Supplementary Table S3). Overall, we largely replicated the matched-group results. However, differences of processing affective ToM emerged in the vmPFC in the matched and full sample in comparison to adults (Supplementary Tables S2 and S3). Taken together, activity of vmPFC seems to vary with age.

DISCUSSION
This study aimed at investigating the neural development of affective ToM processing during adolescence using a dynamic and developmentally sensitive paradigm. Furthermore, we controlled performance via post hoc performance matching. Consistent with previous ToM studies (for a recent meta-analysis, see Mar, 2011), processing the affective ToM film clips across groups resulted in ToM network activation including the vmPFC and dmPFC, the bilateral pSTS/TPJ, the TP, the inferior frontal gyrus, the thalamus and the parahippocampal gyrus. The vmPFC finding is in accordance with both fMRI (Hynes *et al.*, 2006; Sebastian *et al.*, 2012) and lesion studies (Shamay-Tsoory *et al.*, 2006; Shamay-Tsoory and Aharon-Peretz, 2007; Heberlein *et al.*, 2008; Zald and Andreotti, 2010), and the dmPFC finding is in line with Sebastian *et al.* Most importantly, developmental changes in neural activation were observed in the vmPFC with adolescents showing more activation. In contrast, adults activated the right dlPFC more strongly for affective ToM.

Developmental differences in brain activations
Adolescents' stronger activation of the vmPFC for affective ToM
Results show that adolescents had more activation of the vMPFC for affective ToM in contrast to physical control relative to adults. It has been suggested that the vmPFC generates affective meaning (Roy *et al.*, 2012). Specifically, this integrative region recombines complex information from sensory systems, long-term memory and interoceptive cues into future-oriented models of the self and drives decision-making (Roy *et al.*, 2012). This interpretation fits with Shamay-Tsoory *et al.* (2007, 2010), who also describe the vmPFC as a highly integrative region of cognitive and affective information. The current task might require the integration of sensory input by the film clip with past experience of affective ToM states into an affective meaning,

which might facilitate decision making regarding the correct affective state. The observed group difference in vmPFC activity indicates developmental differences in this integration process.

With regard to previous affective ToM studies, our results are in line with Sebastian *et al.* (2012) and Gunther Moor *et al.* (2012) in replicating developmental changes in the vmPFC for affective ToM and are in contrast to Wang *et al.* (2006) because we did not find developmental differences in dmPFC. We could not replicate previous findings of the ongoing development of temporal areas (pSTS, TP; Gunther Moor *et al.* and Wang *et al.*). This might be due to our adolescent participants, who had higher than average affective ToM skills. Most similar to the current Faces test is the Eyes test. Golan *et al.* (2006) showed that the Faces test highly correlated with the Eyes test ($r = 0.74$, $P < 0.01$). In both tests, the participant needs to evaluate facial features of other people and infer the correct affective expression. However, the current task is more realistic since it uses film clips of actors' (i.e. interaction partners') expression of affective ToM.

Adults' stronger modulation of dlPFC resources for affective ToM

We further observed that adults activated the dlPFC more for the affective ToM relative to the physical control condition while adolescents activate this region less for affective ToM relative to physical control. Overall, participants had to keep information in working memory when the stimulus had already disappeared. The dlPFC has been suggested to be implicated in working memory, or more general for monitoring and manipulating cognitive representations (Miller and Cohen, 2001; Koechlin *et al.*, 2003). The demands for working memory between conditions seem to differ. For affective ToM, the information was rather vague: participants needed to form their own impression of the actor's mental state even before the possible choices were shown (which was corroborated by postscanning interviews). In contrast, for physical control, participants knew beforehand the three physical features (colors) that they had to keep in mind. These different demands on working memory seem to modulate dlPFC dependent on age: while adults modulated the dlPFC more in the impression formation of affective ToM, adolescents engaged this region to a greater extent for extracting and remembering of physical features. It is an open question what this developmental pattern means. Future studies could explore if the ongoing development of cognitive resources might lead to the observed pattern.

A limitation of the current study is that the selection of adolescents with same performance as adults might introduce a selection bias. However, if age groups would differ in performance, activation differences in vmPFC might be due to this confound and not actually due to age. Individuals were selected who have higher affective ToM skills than other adolescents with similar age, IQ estimates, socioeconomic background and pubertal status. Therefore, neural age differences between adolescents and adults arise when those individuals, who might develop faster in the affective ToM domain, are included in a sample. Our findings are in line with Sebastian *et al.* (2012). The dlPFC finding appeared robustly in both performance-matched and full sample. Thus, it is age-related, independent of performance between age groups and independent of inter-individual affective ToM differences within the adolescent group. Future studies could investigate how inter-individual differences in affective ToM and age differences interact.

The current fMRI study together with the previous behavioral study (Vetter *et al.*, 2012) follows up on findings of Sebastian *et al.* (2012) in showing ongoing development on both the behavioral and neural level using the same type of affective ToM tasks. Findings suggest that affective ToM continues to develop throughout adolescence. This has been shown using a performance-sensitive task and subsequent control of performance achieved by performance matching. Higher activation in affective ToM *vs* physical control in the vmPFC for adolescents in comparison to adults has been observed. These findings might implicate that adolescents used different neural strategies when performing the task than adults. Overall, one possible reason for the observed functional development might be the prolonged structural development, that is, synaptic pruning of the prefrontal cortex (Giedd, 2008). Specifically, vmPFC and dlPFC undergo gray matter reduction in the course of adolescence (Gogtay *et al.*, 2004; Sowell *et al.*, 2004). Future studies could directly investigate this relationship.

SUPPLEMENTARY DATA

Supplementary Data are available at *SCAN* online.

REFERENCES

Abu-Akel, A., Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, 49, 2971–84.

Ashburner, J., Friston, K.J. (1999). The role of registration and spatial normalisation in detecting activations in functional imaging. *Clinical MRI/Developments in MR*, 7, 26–8.

Bandler, R., Keay, K.A., Floyd, N., Price, J. (2000). Central circuits mediating patterned autonomic activity during active vs. passive emotional coping. *Brain Research Bulletin*, 53, 95–104.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders*, 29, 407–18.

Blakemore, S.J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, 9, 267–77.

Braet, W., Johnson, K.A., Tobin, C.T., et al. (2009). Functional developmental changes underlying response inhibition and error-detection processes. *Neuropsychologia*, 47, 3143–51.

Choudhury, S., Blakemore, S.J., Charman, T. (2006). Social cognitive development during adolescence. *Social Cognitive and Affective Neuroscience*, 1, 165–74.

Church, J.A., Petersen, S.E., Schlaggar, B.L. (2010). The Task B problem and other considerations in developmental functional neuroimaging. *Human Brain Mapping*, 31, 852–62.

Ernst, M., Mueller, S.C. (2008). The adolescent brain: insights from functional neuroimaging research. *Developmental Neurobiology*, 68, 729–43.

Friston, K.J., Holmes, A.P., Poline, J.B., et al. (1995). Analysis of fMRI time-series revisited. *NeuroImage*, 2, 45–53.

Frith, U., Frith, C.D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 358, 459–73.

Galvan, A., Van Leijenhorst, L., McGlennen, K.M. (2012). Considerations for imaging the adolescent brain. *Developmental Cognitive Neuroscience*, 2, 293–302.

Giedd, J.N. (2008). The teen brain: insights from neuroimaging. *Journal of Adolescent Health*, 42, 335–43.

Glaescher, J. (2009). Visualization of group inference data in functional neuroimaging. *Neuroinformatics*, 7, 73–82.

Gogtay, N., Giedd, J.N., Lusk, L., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, 101, 8174–9.

Golan, O., Baron-Cohen, S., Hill, J. (2006). The Cambridge mindreading (CAM) face-voice battery: testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders*, 36, 169–83.

Gunther Moor, B., Op de Macks, Z.A., Güroglu, B., Rombouts, S.A.R.B., Van der Molen, M.W., Crone, E.A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience*, 7, 44–52.

Heberlein, A.S., Padon, A.A., Gillihan, S.J., Farah, M.J., Fellows, L.K. (2008). Ventromedial frontal lobe plays a critical role in facial emotion recognition. *Journal of Cognitive Neuroscience*, 20, 721–33.

Henson, R. (2006). Analysis of fMRI Timeseries: linear time-invariant models, event-related fMRI and optimal experimental design. In: Frackowiak, R.S.J., Friston, K.J., Frith, C., editors. *Human Brain Function*. Oxford: Elsevier Books.

Hynes, C.A., Baird, A.A., Grafton, S.T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia*, 44, 374–83.

Keulers, E.H.H., Evers, E.A.T., Stiers, P., Jolles, J. (2010). Age, sex, and pubertal phase influence mentalizing about emotions and actions in adolescents. *Developmental Neuropsychology*, 35, 555–69.

Koechlin, E., Ody, C., Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302, 1181–5.

Lancaster, J.L., Tordesillas-Gutierrez, D., Martinez, M., et al. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, 28, 1194–205.

Mar, R.A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, 62, 103–34.

Miller, E.K., Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.

Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97–113.

Ollinger, J.M., Shulman, G.L., Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI: I. The method. *NeuroImage*, 13, 210–7.

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press/ Bradford Books.

Petersen, A.C., Crockett, L., Richards, M., Boxer, A. (1988). A self-report measure of pubertal status - reliability, validity, and initial norms. *Journal of Youth and Adolescence*, 17, 117–33.

Price, J.L. (2007). Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. *Annals New York Academy of Sciences*, 1121, 54–71.

Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, 18, 2188–99.

Roy, M., Shohamy, D., Wager, T.D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, 16, 147–56.

Saxe, R., Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, 19, 1835–42.

Schlaggar, B.L., Brown, T.T., Lugar, H.M., Visscher, K.M., Miezin, F.M., Petersen, S.E. (2002). Functional neuroanatomical differences between adults and school-age children in the processing of single words. *Science*, 296, 1476–9.

Sebastian, C.L., Fontaine, N.M., Bird, G., et al. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, 7, 53–63.

Serences, J.T. (2004). A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. *NeuroImage*, 21, 1690–700.

Shamay-Tsoory, S.G., Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, 45, 3054–67.

Shamay-Tsoory, S.G., Harari, H., Aharon-Peretz, J., Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, 46, 668–77.

Shamay-Tsoory, S.G., Tibi-Elhanany, Y., Aharon-Peretz, J. (2006). The ventromedial prefrontal cortex is involved in understanding affective but not cognitive theory of mind stories. *Social Neuroscience*, 1, 149–66.

Sowell, E.R., Thompson, P.M., Toga, A.W. (2004). Mapping changes in the human cortex throughout the span of life. *Neuroscientist*, 10, 372–92.

Steinberg, L., Morris, A.S. (2001). Adolescent development. *Annual Review of Psychology*, 52, 83–110.

Stone, V.E., Baron-Cohen, S., Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10, 640–56.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, 30, 829–58.

Vetter, N.C., Altgassen, M., Phillips, L., Mahy, C.E.V., Kliegel, M. (2013). Development of affective theory of mind across adolescence: disentangling the role of executive functions. *Developmental Neuropsychology*, 38, 114–25.

Vetter, N.C., Leipold, K., Kliegel, M., Phillips, L.H., Altgassen, M. (2012). Ongoing development of social cognition in adolescence. *Child Neuropsychology*, doi: 10.1080/ 09297049.2012.718324 [Epub ahead of print].

Völlm, B.A., Taylor, A.N.W., Richardson, P., et al. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, 29, 90–8.

Wang, A.T., Lee, S.S., Sigman, M., Dapretto, M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive and Affective Neuroscience*, 1, 107–21.

Watzlawik, M. (2009). Die Erfassung des Pubertätsstatus anhand der Pubertal Development Scale. *Diagnostica*, 55, 55–65.

Zald, D.H., Andreotti, C. (2010). Neuropsychological assessment of the orbital and ventromedial prefrontal cortex. *Neuropsychologia*, 48, 3377–91.