

Fairness decisions in response to emotions: a functional MRI study among criminal justice-involved boys with conduct disorder

Eduard T. Klapwijk,^{1,2} Gert-Jan Lelieveld,^{2,3} Moji Aghajani,^{1,2} Albert E. Boon,^{1,4,5} Nic J. A. van der Wee,^{2,6} Arne Popma,^{7,8} Robert R. J. M. Vermeiren,^{1,2,7} and Olivier F. Colins^{1,2}

¹Department of Child and Adolescent Psychiatry, Curium—Leiden University Medical Center, Leiden, The Netherlands, ²Leiden Institute for Brain and Cognition (LIBC), Leiden, The Netherlands, ³Institute of Psychology, Leiden University, Leiden, The Netherlands, ⁴Scientific Research Department, Stichting De Jutters, The Hague, The Netherlands, ⁵Lucertis, Centre for Child and Adolescent Psychiatry, Rotterdam, The Netherlands, ⁶Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands, ⁷Department of Child and Adolescent Psychiatry, VU University Medical Center, Amsterdam, The Netherlands, and ⁸Institute of Criminal Law & Criminology, Faculty of Law, Leiden University, Leiden, The Netherlands

Correspondence should be addressed to Eduard T. Klapwijk, Curium-LUMC, Postbus 15, 2300 AA Leiden, The Netherlands. E-mail: e.t.klapwijk@curium.nl, et.klapwijk@gmail.com.

Abstract

Research suggests that individuals with conduct disorder (CD) are marked by social impairments, such as difficulties in processing the affective reactions of others. Little is known, though, about how they make decisions during social interactions in response to emotional expressions of others. In this study, we therefore investigated the neural mechanisms underlying fairness decisions in response to communicated emotions of others in aggressive, criminal justice-involved boys with CD ($N = 32$) compared with typically developing (TD) boys ($N = 33$), aged 15–19 years. Participants received written emotional responses (angry, disappointed or happy) from peers in response to a previous offer and then had to make fairness decisions in a version of the Dictator Game. Behavioral results showed that CD boys did not make differential fairness decisions in response to the emotions, whereas the TD boys did show a differentiation and also responded more unfair to happy reactions than the CD boys. Neuroimaging results revealed that when receiving happy vs disappointed and angry reactions, the CD boys showed less activation than the TD boys in the temporoparietal junction and supramarginal gyrus, regions involved in perspective taking and attention. These results suggest that boys with CD have difficulties with processing explicit emotional cues from others on behavioral and neural levels.

Key words: social decision-making; conduct disorder; callous-unemotional; Dictator Game; fMRI

Introduction

Individuals with conduct disorder (CD) are characterized by a persistent pattern of aggressive and antisocial behavior

(American Psychiatric Association, 2013), along with marked socioemotional deficits and interpersonal difficulties (Dodge, 1993; Happe and Frith, 1996; Schwenck *et al.*, 2012). These

Received: 13 July 2015; Revised: 23 November 2015; Accepted: 11 December 2015

© The Author (2016). Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

socioemotional and interpersonal deficits are expressed in reduced responses to the distress cues of others and lack of care about others' suffering, especially in individuals with CD who show elevated levels of callous-unemotional (CU) traits¹ (Pardini, 2011; Blair, 2013; Lockwood et al., 2013). Studies that investigated how brain regions involved in social cognition function differently in CD (regardless of the level of CU traits) have mainly used static stimuli such as pictures of emotional faces or scenarios (e.g. Herpertz et al., 2008; Marsh et al., 2013) or stories about mental states (e.g. Sebastian et al., 2012). Although these studies have greatly increased our understanding of the neurocognitive abnormalities in processing social stimuli in youth with CD (for a review see Blair, 2013), most do not take into account the interactive nature of social exchange, which is one of the hallmarks of social interaction. Yet social neuroscientists recently started to use simple but sophisticated tasks derived from experimental economics to study social decision-making in an interactive context (Rilling and Sanfey, 2011) and to study aberrant social decision-making in clinical populations (Kishida et al., 2010; Hasler, 2012). These tasks can be used to study a range of behaviors such as trust, fairness, altruism and social norm compliance, which might in turn be influenced by individual variations in personality traits such as empathy.

More specifically, several studies have used economic games to examine social decision-making in relation to antisocial behavior and psychopathic traits in adults and adolescents. Neuroimaging studies showed that psychopathic traits in adults are positively related to uncooperative behavior in economic games and to weaker responses in brain regions important for processing social cues, such as the orbitofrontal cortex and the amygdala (Rilling et al., 2007; Mokros et al., 2008; Koenigs et al., 2010). One study examined the influence of reputations of others during a social exchange game and found that youths with externalizing behavior problems compared with typically developing (TD) youth show reduced differential responses within the anterior insula and caudate to the offers of a neutral relative to a kind or an aggressive partner (Sharp et al., 2011). Another neuroimaging study showed that criminal justice-involved boys were less willing to accept lower offers from others compared with TD boys, even if they knew the other had no choice (van den Bos et al., 2014). In these criminal justice-involved boys, higher callousness scores were also related to fewer acceptances when the other had no choice compared to when the other had a fair alternative. This was accompanied by less activity in the right temporoparietal junction (rTPJ), a brain region important for social cognition and attention (Van Overwalle and Baetens, 2009; Krall et al., 2015). The TPJ appears to be a site of convergence for social and attention processing streams, in which social context is extracted and synthesized in order to guide attention and decision-making (Carter and Huettel, 2013). These results suggest that the criminal justice-involved boys were mainly focused on the unfairness of the offers and less influenced by the perspective of the other player (van den Bos et al., 2014). Altogether, economic game studies show that antisocial individuals are less inclined than healthy individuals to take contextual information into account during social exchanges (Sharp et al., 2011; Radke et al., 2013; van den Bos et al., 2014).

¹ Callous-unemotional (CU) traits are a circumscribed facet of psychopathy and refer to a set of affective features characterized by deficient empathy and guilt, insensitivity to others' feelings, and shallow emotions.

In contrast, evidence from healthy populations shows that contextual information in the form of emotions expressed by others heavily influence social decisions (van Kleef et al., 2010). For example, people react with more fair offers after they read disappointed compared with angry reactions, probably due to feelings of guilt caused by disappointment (Lelieveld et al., 2012, 2013b). In addition, higher psychopathic trait-scores in undergraduate students were found to be related to a lack of response to emotional feedback of happiness in an economic game (Johnston et al., 2014). To date, no study that used an interactive economic game in antisocial populations focused on the role of other's emotions in social interactions. Although individuals with CD (and especially those with high CU traits) are known to have problems with processing the affective reactions of others (Jones et al., 2009; Schwenck et al., 2012; Sebastian et al., 2012), little is known about how they make social decisions in response to emotions in an interactive context.

In this study, we therefore investigated the effects of other's emotions on fairness decisions and associated brain responses in boys with CD compared with TD controls. Participants had to allocate tokens between themselves and peers from which they received verbal emotional reactions depicting anger, disappointment or happiness (Lelieveld et al., 2013a). This procedure allowed us to test whether boys with CD would differentiate between various emotions and would adjust their fairness decisions accordingly. A behavioral study that used this paradigm found that TD adolescents took emotional reactions of others into account and reacted with more fair offers after they read disappointed reactions compared with angry and happy reactions from their peers (Klapwijk et al., 2013). In addition, in a neuroimaging study that used this paradigm healthy adults showed more activation in the rTPJ when receiving happy reactions (and they reacted with more fairness in response to both happy and disappointed reactions compared with angry reactions), suggesting increased perspective taking and attention in response to happiness (Lelieveld et al., 2013a). Based on studies pointing to problems in processing affective and contextual social signals of both negative and positive emotions in CD (Herpertz et al., 2005; Fairchild et al., 2009; de Wied et al., 2012), we expected that the CD boys would be less responsive to emotional information of others. Such low emotional responsiveness might lead to a decrease in differentiating between emotions. We expected that this lower emotional responsiveness would be reflected in less differentiation in fairness decisions between the three emotions in the CD (*vs* TD) boys, and by less activation in social-cognitive brain areas such as the TPJ and medial prefrontal cortex (MPFC) in the CD (*vs* TD) boys. Additionally, we investigated the effects of CU traits on brain and behavior in our task. Based on prior work, it was hypothesized that the CD boys with high CU traits would show even more difficulties in differentiating between negative and positive emotions than CD boys with low CU traits (de Wied et al., 2012; Fanti et al., 2016).

Materials and Methods

Participants

Since CD is highly prevalent among criminally justice-involved boys (Colins et al., 2010), adolescent offenders with CD were recruited from a juvenile detention center and a forensic psychiatric facility. All had been convicted or charged for felony crimes such as assault, murder or armed robbery. Typically developing (TD) control adolescents were recruited through local

advertisement. All participants were aged 15–19 years (Table 1 for participant characteristics). Exclusion criteria for all participants were (central) neurological abnormalities, a history of epilepsy or seizures, head trauma, left-handedness and $IQ < 75$. Data from participants with excess motion defined by relative mean displacement > 0.5 mm were excluded from further analysis. Of note, the current task was part of a larger study and preceded by other scans [e.g. structural magnetic resonance imaging (MRI), resting state functional MRI (fMRI)], which might have increased the likelihood of excessive head motion during this task. To obtain an estimate of intelligence, participants completed the Wechsler Adult Intelligence Scale—third edition (WAIS-III) or Wechsler Intelligence Scale for Children—third edition (WISC-III) subscales Vocabulary and Block Design. CU traits were measured using the Inventory of Callous-Unemotional traits (ICU; Kimonis et al., 2008).

The CD group consisted of 54 adolescent boys of which 46 completed both phases of the experimental fMRI task (see Experimental task section below). Diagnoses were confirmed using the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS-PL) Behavioral Disorders screening (Kaufman et al., 1997), a widely used semi-structured diagnostic interview. Only boys who fulfilled DSM-IV-TR criteria for CD with at least one aggressive symptom (e.g. used a weapon, has been physically cruel to people, has stolen while confronting a victim) were included. Data from 14 CD participants were discarded due to excessive motion, leaving a final sample of 32 participants with CD. The excluded CD participants did not significantly differ from the CD participants that were included in the fMRI analysis in age, comorbidity, callous-unemotional traits, BES affective and cognitive empathy scores or unfairness percentages in response to the three emotions in the experimental task (all $P_s > 0.2$). The groups did differ in estimated IQ scores ($P < 0.005$), caused by lower IQ scores in the excluded (92.1) vs the included group (98.1). Eight participants with CD also met DSM-IV-TR criteria for ADHD. No other comorbid disorders were reported and none of the participants with CD took medication at the time of testing (medication history was not recorded).

Thirty-seven TD control boys were recruited through local advertisement of which 34 completed both phases of the task (see Experimental task section below). These participants were screened using the K-SADS-PL Behavioral Disorders module in order to exclude participants with behavioral disorders. The Youth Self Report (YSR; Achenbach, 1991) was used to assess general psychopathology; none of the TD boys scored in the

clinical range on the YSR externalizing and internalizing scales. Data from one TD participant were discarded due to excessive motion, leaving a final sample of 33 TD participants. The CD group showed more head motion than the TD group and we therefore had to exclude more CD than TD participants. Importantly, in the final sample used in our paper there is no difference in relative mean displacement ($P > 0.19$) between the CD and TD groups.

Experimental task

We examined participants' fairness choices in the Dictator Game (Kahneman et al., 1986; Güroğlu et al., 2009) after receiving emotional reactions from others, using a procedure previously used in studies with adults and adolescents (Klapwijk et al., 2013; Lelieveld et al., 2013a). One week before participants took part in the scanning session, they first participated in a preliminary study (first phase of the experiment). This phase was used to create an interpersonal context for the emotional reaction they later (second phase) received. In the first phase, participants read a scenario after which they were instructed to divide 10 tokens between themselves and another person. They could choose a 6–4 distribution in favor of themselves, an equal distribution (5–5) or a distribution in favor of the other (4–6). This negotiation scenario was intended to assure that most participants chose the 6–4 option in this phase of the study. Only participants that chose a 6–4 distribution took part in the second phase of the experiment during scanning (46 out of 54 CD boys and 34 out of 37 TD boys chose a 6–4 distribution). This was done to ensure credibility of the second phase in which emotional reactions would be directed at the 6–4 offer chosen in the first phase. In line with previous studies (van Kleef et al., 2010; Lelieveld et al., 2013a), these reactions were angry, disappointed or happy. Using these three emotions allows for comparisons of the effects of negative and positive communicated emotions and the effects of different types of negative emotions. Additionally, although it is not uncommon to find angry and disappointed reactions in response to a 6–4 distribution because of the relative unfairness of this distribution, happy reactions should be considered acceptable since offers of around 40% of the total are mostly accepted in economic games (Falk and Fischbacher, 2006).

In the second phase of the experiment, the boys were told that their unfair offer (the 6–4 distribution chosen in the first phase) was presented to 60 peers who were given the opportunity to write out their reaction upon receiving the offer. In reality, the reactions were preprogrammed and we left at least 1 week between the first and second phase to increase the credibility that researchers actually collected reactions from others. During scanning (also part of the second phase), participants were paired with a different player on each trial, whose first name was provided and whose reaction to the 6–4 distribution was angry, disappointed or happy. These preprogrammed reactions were rated to reflect the intended emotion (see also Klapwijk et al., 2013; Lelieveld et al., 2013a). Participants read the reactions of their peers and subsequently played a version of the Dictator Game with the peer who provided the reaction (Figure 1). In this Dictator Game the participants were the allocator and had to divide 10 tokens. They could now choose between different fair and unfair distributions and learned that the recipient had to accept any distribution they would make. The possible distributions were 5–5 vs 7–3; 6–4 vs 4–6; 3–7 vs 7–3; and 5–5 vs 6–4; and all options were presented five times during each emotion type. Each trial started with a jittered fixation

Table 1. Participant characteristics

	Conduct disorder (CD) (N = 32)	Typically developing (TD) (N = 33)
Age, years (s.d.)	16.8 (1.2)	17.2 (1.2)
IQ, M (s.d.)	98.1 (7.0)	97.2 (8.7)
Minority, N (%)	27 (84.4)	9 (27.3)
Empathy scores ^a		
Cognitive empathy, M (s.d.)	36.3 (5.9)	38.0 (5.0)
Affective empathy, M (s.d.)**	28.9 (7.7)	36.1 (7.8)
Callous-unemotional traits, M (s.d.)*	26.0 (11.2)	20.8 (7.1)

^aSelf-report of affective and cognitive empathy was measured using the Basic Empathy Scale (Jolliffe and Farrington, 2006).

*Significantly different at $P < 0.05$.

**Significantly different at $P < 0.001$.

(min. = 0.55 s, max. = 4.95 s, $M = 1.54$ s), after which the participants were presented with the emotional reaction for a period of three seconds plus a jittered interval (min. = 0.55 s, max. = 4.95 s, $M = 1.54$ s) and subsequently had 6 s to make a decision between two distributions. The 60 trials were presented in pseudo-random order divided over three blocks of 4 min each. Before the task started, participants learned that at the end of the experiment the computer would randomly select 10 trials to determine their total earnings, which would be added to the standard compensation for their participation. At the end of the session, participant's pay-off was presented, which varied between 2.5 and 6 euros. Afterwards, participants completed a post-scanning questionnaire in which they were probed for suspicion and asked to indicate their levels of guilt, anger and fear in response to the different emotions. None of the participants expressed doubt about the set-up of the task.

fMRI data acquisition

Imaging was carried out at the Leiden University Medical Center on a 3T Philips Achieva MRI scanner. Prior to scanning, participants were familiarized with the scanner environment using a mock scanner. For fMRI, T2* weighted gradient echo, echo planar images sensitive to blood oxygen level-dependent (BOLD) contrast were obtained with the following acquisition parameters: repetition time (TR) = 2.2 s, echo time (TE) = 30 ms, flip angle = 80°, 38 axial slices, field of view (FOV) = 220 × 220 mm, 2.75 mm isotropic voxels, 0.25 mm slice gap. A high-resolution anatomical image (T₁-weighted ultra-fast gradient-echo acquisition; TR = 9.75 ms, TE = 4.59 ms, flip angle = 8°, 140 axial slices, FOV = 224 × 224 mm, in-plane resolution 0.875 × 0.875 mm, slice thickness = 1.2 mm) was acquired for registration purposes. All anatomical scans were reviewed by a radiologist; no anomalies were found.

fMRI data analysis

fMRI data analysis was conducted using FEAT (fMRI Expert Analysis Tool) version 6.00, part of FSL (www.fmrib.ox.ac.uk/fsl). Data pre-processing consisted of motion correction using MCFLIRT, non-brain removal using BET, spatial smoothing using a Gaussian kernel of full width at half-maximum (FWHM) 5 mm, grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with $\sigma = 50.0$ s). Functional scans were registered to

the T1-weighted images, and subsequently to the 2 mm MNI-152 standard space template. Time-series statistical analysis was performed using FILM with local autocorrelation correction. To investigate the effects of the communicated emotions, we modeled the onset of the presentation of the three different emotional reactions (i.e. anger, disappointment, happiness) as an event with zero duration convolved with a gamma hemodynamic response function. To account for residual movement artifacts, the six realignment parameters were included in the model as covariates of no interest. At first-level for each run for each participant, primary contrasts of interest were generated. Positive vs negative emotions were contrasted (happiness > [anger and disappointment]) as well as happiness against the separate negative emotions (happiness > anger; happiness > disappointment) and the negative emotions against each other (anger > disappointment). A second-level, fixed-effects analysis combined data across the three runs for each participant. Individual participant data were then entered into a third-level group analysis using a mixed-effects design (FLAME) whole-brain analysis. The general linear model included the two groups (CD and TD) and to account for possible age effects, we included age (mean-centered) as covariate of no interest. Resulting statistical maps were corrected for multiple comparisons using cluster-based correction ($P < 0.05$, initial cluster-forming threshold $z > 2.3$). We used Featquery and SPSS to conduct region of interest (ROI) analyses to correlate task behavior and ICU scores with patterns of activity from regions that were identified in the whole-brain analyses. Functional ROIs from these regions were generated by masking the activation maps of the contrasts of interest with binarized anatomical ROIs using the Harvard-Oxford structural atlases distributed with FSL. Finally, we explored whether comorbid ADHD in the CD group might have influenced the results. Extracted z values from the ROIs identified in the whole-brain analyses were entered into SPSS to compare only those participants with CD without comorbid ADHD to TD controls and to compare boys with and without comorbid ADHD with each other.

Results

Behavioral results

Fairness decisions after the three different emotions were compared between the groups with a 2 × 3 mixed ANOVA (group × emotion). We found a main effect of emotion, $F(1, 64) = 8.47$, $P = 0.001$, caused by a higher percentage of unfair offers in

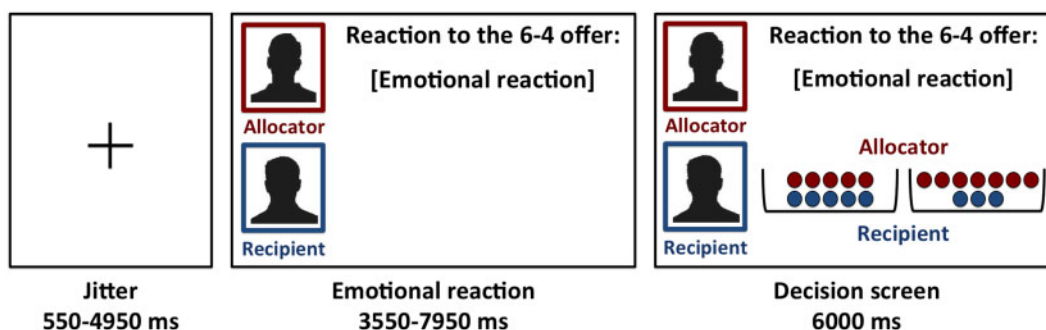


Fig. 1. Visual display and timing (in milliseconds; ms) of the task in the scanner. The emotional reaction of the recipient (here 'emotional reaction') was displayed after a jittered fixation cross. Subsequently, the screen displayed two offers each containing red and blue tokens, which indicated the share for the allocator and the recipient, respectively (here 5-5 vs 7-3). The name of the allocator was displayed in red (here 'allocator') and the name of the recipient in blue (here 'recipient'). If participants did not respond within 6000 ms, a screen displaying "Too late!" was presented. After the response, the decision screen remained on the screen until 6000 ms after the onset of the decision screen.

response to angry ($M = 56.4\%$; $s.d. = 32.9$) compared with disappointed reactions ($M = 48.4\%$; $s.d. = 30.0$, $P = 0.001$). We found no main effect of group, $F(1, 64) = 2.75$, $P = 0.102$, showing that the groups did not differ on fairness levels across the emotions combined. The interaction effect was trendwise significant, $F(1, 64) = 2.62$, $P = 0.081$, indicating group differences in the reactions after the different emotional expressions. Analyses of the CD and TD participants separately revealed that the CD participants made no difference in fairness decisions after reading the different emotions, $F(2, 64) = 1.21$, $P = 0.31$, whereas the TD participants did, $F(2, 66) = 11.66$, $P < 0.001$. In line with Klapwijk et al. (2013), *post hoc* tests revealed that TD participants more often chose the unfair than the fair option when dealing with angry recipients (59.5%, $s.d. = 33.1$, $P < 0.001$) and happy recipients (66.8%, $s.d. = 27.1$, $P < 0.05$) than when dealing with disappointed recipients (49.2%, $s.d. = 31.1$). The percentage of unfair offers in response to happy and angry recipients ($P = 0.36$) did not differ in the TD group. Thus, communications of disappointment elicited relatively more fair offers than communications of anger and happiness did, but only in the TD and not in the CD group (Figure 2). Finally, between-group comparisons showed that the TD group made more unfair offers after happy ($P = 0.005$) but not after angry ($P = 0.83$) or disappointed ($P = 0.45$) reactions than the CD group.

Correlations between post-scanning ratings (guilt, anger, fear) and fairness decisions revealed that self-reported guilt when reading angry reactions correlated negatively with unfair offers in response to angry reactions in the TD group ($r = -0.54$, $P < 0.001$), but not in the CD group ($r = -0.31$, $P = 0.10$), and that self-reported guilt after disappointed reactions correlated negatively with unfair offers in response to disappointment in the TD group ($r = -0.52$, $P < 0.005$), but not in the CD group ($r = -0.20$, $P = 0.30$). Fisher z -values were calculated to compare the correlations between the CD and TD groups. No significant group difference was found for the correlation between self-reported guilt and unfair offers in response to anger ($z = 1.09$, $P = 0.14$) and a trendwise significant difference was found for the correlation between self-reported guilt and unfair offers in response to disappointment ($z = 1.43$, $P = 0.076$). These results suggest that levels of guilt in the TD control were associated with individual differences in fairness decisions in reaction to disappointed reactions, whereas no significant relation was found for the CD group.

To further explore the role of CU traits in the CD group, we also conducted an analysis in which we separated the CD group into a group with high CU traits (CD/CU+; $N = 14$) and a group with low CU traits (CD/CU-; $N = 18$). Participants scoring above the median ICU score of the full CD sample ($N = 54$; median score = 27.0) were included in the CD/CU+ group and those

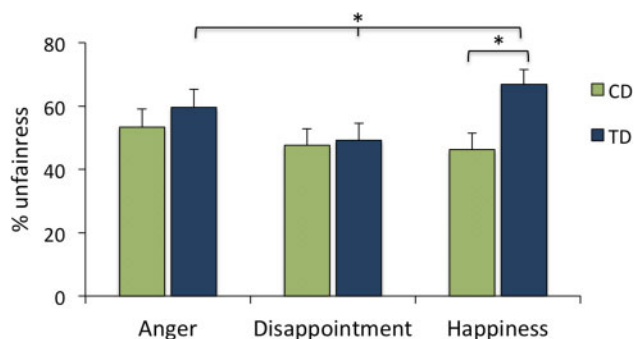


Fig. 2. Percentage of unfair offers after communication of anger, disappointment and happiness, separate for CD and TD groups.

scoring on or under the median ICU score in the CD/CU- group. This analysis did not reveal differences between the CD/CU+ and CD/CU- group on behavior; both groups made no differences in fairness decision between the three emotions, $F(2, 28) = 0.49$, $P = 0.62$ (CD/CU+), and, $F(2, 36) = 0.66$, $P = 0.53$ (CD/CU-).

fMRI results

The first set of whole-brain analyses investigated regions that showed group differences between the CD and TD groups when receiving positive relative to negative emotional reactions in general (i.e. happiness > [anger and disappointment] contrast). This analysis revealed that the CD group showed less activation than the TD group in a cluster in the rTPJ and right supramarginal gyrus (rSMG) (Figure 3A and B), a cluster in the left superior parietal lobule and a cluster in the somatosensory cortex (Table 2). No regions were found where the CD group showed more activation than the TD group in this contrast. When analyzing the contrasts that compared happiness to a specific negative emotion (i.e. happiness > anger, and the happiness > disappointment), group differences remained in the rSMG. Furthermore, in the happiness > anger contrast, we also found less activation in the right dorsolateral prefrontal cortex (rDLPFC, Figure 4A and B) in the CD compared with the TD group. Finally, when comparing the two negative emotions with each other, we found no significant group differences between the CD and TD groups when analyzing the anger > disappointment and disappointment > anger contrasts. Additionally, we re-analyzed the fMRI data using a stricter cluster-corrected threshold of $z > 3.1$, $P < 0.05$, after which the group differences of the whole brain comparisons were not significant anymore.

Relationships between fairness decisions and brain activation

Next, we conducted exploratory analyses to investigate the relation between fairness decisions and brain activity in regions identified in our whole-brain analysis. Because of the differences found in the happy condition between the CD and TD groups, these analyses focused on the behavioral and brain responses during the happy condition. We investigated the relation between the percentage of unfair offers in response to happy reactions and the activity in the rTPJ/SMG for the happy > fixation contrast. This analysis revealed a significant negative correlation between the percentage unfair offers and rTPJ/SMG activity for the TD control group ($r = -0.55$, $P < 0.001$), but not for the CD group ($r = 0.29$, $P = 0.11$, Figure 3C). Additionally, Fisher z -values were calculated which indicated that the correlations differed significantly between the groups ($z = -3.52$, $P < 0.001$). Thus, TD boys who showed higher (vs lower) levels of rTPJ/SMG activation when happiness was expressed tended to react more fair after happy reactions. This latter finding demonstrates that for the TD control group rTPJ/SMG activation is associated with individual differences in fairness decisions in reaction to happy reactions, whereas no significant relation was found for the CD group.

In addition, the relation between the percentage of unfair offers in response to happy reactions and the activity in the rDLPFC for the happy > fixation contrast revealed a significant negative correlation between the percentage unfair offers and rDLPFC activity for the TD control group ($r = -0.41$, $P < 0.05$), but not for the CD group ($r = -0.20$, $P = 0.28$, Figure 4C). However, Fisher z -values were calculated which indicated that these

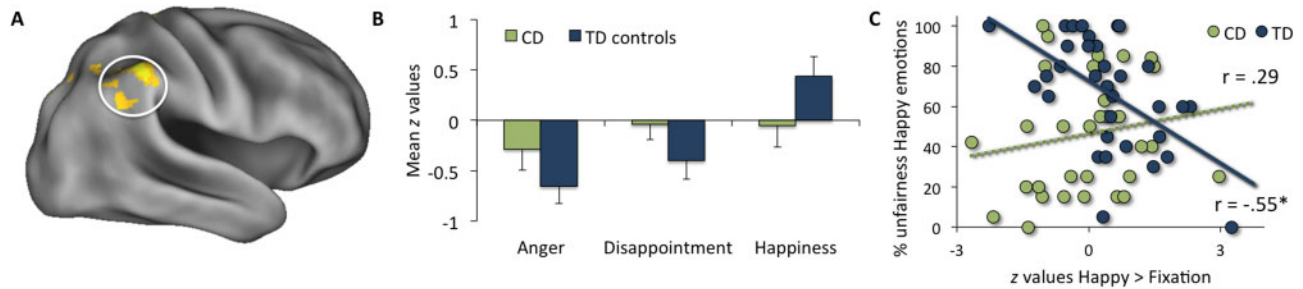


Fig. 3. (A) rTPJ/SMG group differences in the happiness > [anger and disappointment] contrast cluster-thresholded at $z > 2.3$, $P < 0.05$ with (B) mean z values plotted for the three emotions and the CD and TD groups separately. (C) Activation in the rTPJ/SMG in the [happy > fixation] condition correlated negatively with the percentage unfair offers in response to happy emotions for the TD control group, but not for the CD group. Fisher z-values indicated that the correlations differed significantly between the groups ($z = -3.52$, $P < 0.001$).

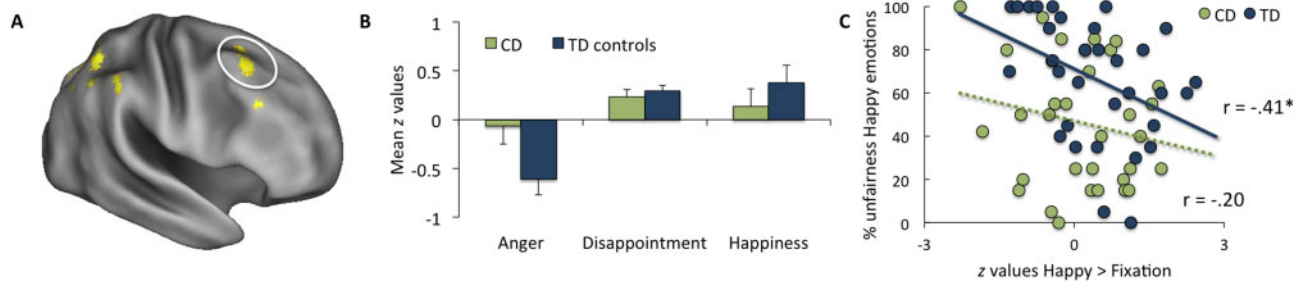


Fig. 4. (A) rDLPFC group differences in the happiness > anger contrast cluster-thresholded at $z > 2.3$, $P < 0.05$ with (B) mean z values plotted for the three emotions and the CD and TD groups separately. (C) Activation in the rDLPFC in the [happy > fixation] condition correlated negatively with the percentage unfair offers in response to happy emotions for the TD control group, but not for the CD group. However, Fisher z-values indicated that these correlations did not differ significantly between the groups ($z = -0.89$, $P = 0.19$).

Table 2. Montreal Neurological Institute (MNI) coordinates, z values and cluster size for brain regions revealed by the whole-brain pairwise comparisons of the TD control > CD groups, $z > 2.3$, $P < 0.05$ cluster-corrected

Anatomical region	Max z	MNI peak coordinates			Size in voxels
		x	y	z	
Happiness > [anger and disappointment]					
R supramarginal gyrus extending to the temporoparietal junction	4.26	58	-38	50	1064
Precentral gyrus	4.26	-2	-32	66	605
L superior parietal lobule	4.24	-40	-56	60	1801
Happiness > anger					
L superior parietal lobule	4.40	-40	-56	60	1428
R middle frontal gyrus (DLPFC)	4.19	38	22	52	439
R supramarginal gyrus	3.75	58	-38	52	1072
Happiness > disappointment					
R supramarginal gyrus	4.2	56	-36	48	750

Notes: Activation clusters were labeled using the Harvard–Oxford structural atlases.

correlations did not differ significantly between the groups ($z = -0.89$, $P = 0.19$).

Effects of CU traits on brain activation

No significant relation between brain activation in ROIs derived from the whole-brain analysis and variation of CU traits were found within the CD group or within the TD group. To further explore the role of CU traits in the CD group, we also conducted analyses with the CD/CU+ ($N = 14$) and CD/CU- ($N = 18$) groups separately (see Behavioral results). These analyses did not reveal any significant group differences between the CD/CU+ and CD/CU- groups.

Effects of comorbidity

Post-hoc analyses revealed that all group differences remained significant when excluding CD boys with comorbid ADHD (all P s < 0.001). In addition, no significant group differences were found between CD participants with comorbid ADHD and those without (all P s > 0.3).

Discussion

The current study investigated behavioral and neural responses in reaction to other's emotions in an interactive context in CD and TD boys. Behavioral results suggest that the CD boys differentiate less between different emotions communicated by

others when making fairness decisions than TD boys. In line with prior work with TD adolescents (Klapwijk et al., 2013), TD boys reacted relatively more fair in response to disappointed reactions compared with angry and happy reactions, whereas the CD boys did not show differences in fairness reactions between the three emotions. These results are in line with previous studies that suggest that individuals with CD have difficulties in processing affective stimuli (Herpertz et al., 2005; Fairchild et al., 2009), and our study contributes to the literature by showing that CD boys do not adjust their allocation behavior in response to emotional information of others. On the other hand, one might have expected that the CD boys would react more unfair in response to anger since they are more easily provoked by angry reactions and pay more attention to hostile cues (Dodge, 1993). However, we found no differences between CD boys and TD boys in fairness decision in response to angry reactions. Since hostile attributions are mostly focused on ambiguous content (de Castro et al., 2002), we might have found an influence of hostile attribution of intent in the CD boys had we used more ambiguous instead of clearly angry reactions. The current results, nevertheless, can be interpreted as a sign of insensitivity to emotions in the CD group reflected by equal amounts of fairness in response to different emotions.

The fMRI results showed that the CD boys compared with the TD boys had less activity in the right rTPJ/SMG when receiving happy compared with disappointed and angry reactions. This is in line with a previous study using this paradigm that reported increased rTPJ activation in healthy adults in this contrast (Lelieveld et al., 2013a). The rTPJ and also the nearby-located rSMG are important regions for social cognitive abilities such as perspective taking and empathy (Frith and Frith, 2006; Silani et al., 2013; Krall et al., 2015). Thus, based on these prior studies, the decreased activation in these brain areas in the CD group might suggest that boys with CD were less inclined to take the perspective of the other person during happy compared with angry and disappointed reactions. In the current paradigm, the TPJ might support the integration of information streams to construct a social context, which may then be used to adapt behavioral decisions in response to other's emotions (cf., Carter and Huettel, 2013). Additionally, the negative correlation between right rTPJ/SMG activation and unfairness in response to happy reactions that we found only in TD controls suggests that taking the perspective of the other resulted in less unfair offers in the TD but not the CD boys. Hence, this correlational analysis supports the idea that the TD boys are more sensitive to the emotions of others than the CD boys and consequently adapt their behavior in response to others' emotions. It should be noted, however, that although activation in the rTPJ/SMG was associated with more fair offers after happiness in the TD and not the CD group, the TD group made more unfair offers in response to happiness than the CD group.

Although we hypothesized decreased activation in various brain regions involved in social processing such as the TPJ and MPFC, in the current study group differences seem to be selective for TPJ/SMG. However, we also found decreased rDLPFC activation in the CD compared with the TD group when reading happy vs angry reactions. The rDLPFC is an important region implicated in cognitive control (Miller and Cohen, 2001) and plays a role in regulating reactions and implementing norm compliance in social decision-making (Spitzer et al., 2007; Rilling and Sanfey, 2011; Steinbeis et al., 2012). Decreased activation in this area might be suggestive of less regulatory brain activation in the CD boys compared with controls. The negative association between rDLPFC activation and unfair decisions in

response to happy reactions in the TD control group suggests that withholding the urge to make an unfair decision requires cognitive control. However, one may then also expect more unfair offers in response to happiness in the CD vs TD group, which was not the case in the current study. Therefore, a likely alternative explanation is that in line with the important role of rDLPFC, rTPJ and rSMG in attentional processes (Corbetta et al., 2008; Mitchell, 2008; Ochsner et al., 2012), reduced activation in these areas might reflect reduced attention to the happy expressions in the CD vs TD boys.

Many studies have found that impaired emotional responsiveness in adolescents with conduct problems or with CD is more pronounced in those with high CU traits (Frick et al., 2014). However, current results suggest that boys with CD have difficulties in adapting their behavior in response to emotions of others irrespective of whether they show elevated levels of CU traits. This is consistent with some previous work showing non-social decision-making deficits in antisocial youth irrespective of CU traits (White et al., 2014). In addition, previous studies that did find effects of CU traits on emotional responsiveness have mostly used facial emotions in which most effects were found for distress cues such as fear and sadness (but see Dawel et al., 2012 for meta-analytic evidence for more broad impairments in emotion recognition). It might be that responsiveness in the form of social decisions to written emotional reactions that depict anger, disappointment and happiness as employed in our task might not be associated with CU traits.

Some limitations of the current study should be considered. Although the study design focused on the effects of different emotions on fairness decisions and not on fairness per se, it must be noted that although the groups did not differ on total unfairness across the three emotions, contrary to what one might expect the CD group behaved less unfair in response to happy reactions than the TD controls. However, higher unfairness in the Dictator Game in controls vs adult inmates has been reported previously and has been interpreted as a form of compensation to amend for their crimes (Gummerum and Hanoch, 2012). The criminal justice-involved CD boys might also have been motivated by a desire to please the experimenters, if they thought that despite guaranteed anonymity their behavior would be reported to authorities. Nevertheless, being more sensitive to the different emotions, the TD participants could have concluded that the happy other was satisfied with the previous unfair offer, and therefore would be content with another unfair offer (van Kleef et al., 2010). Another possible caveat of the current study is that our design does not allow for inferences about whether the equal distributions in response to different emotions in the CD group reflect less differentiation between emotions or that this reflects that CD youth just do not use the emotional information when making fairness decisions. However, the relation between feelings of guilt (as reported after the scanning procedure) and fairness in response to disappointment in the TD controls but not in the CD group suggest that feelings of guilt did not influence fairness decisions after disappointment in the CD group. Future studies are needed in which emotion states or skin conductance are being measured directly when CD boys read the emotions in order to answer whether the differentiation is indeed being hampered as a consequence of less emotional responsiveness. Another limitation of the current study is that our group differences are reported at a cluster-corrected threshold of $z > 2.3$, $P < 0.05$. Notwithstanding that this is a widely used correction method, it can result in false positives and low spatial specificity (Woo et al., 2014). Using a stricter cluster-corrected threshold of $z > 3.1$, $P < 0.05$, as

suggested by Woo et al. (2014), the group differences in the present study did not remain significant. However, when studying social-affective processes in difficult to recruit detained male adolescents, we should also be careful and avoid false negatives [see recommendations from Lieberman and Cunningham (2009)]. Our study is a first step in examining how explicit emotional feedback influences brain and behavior in criminal justice-involved CD boys and future studies are needed to replicate these findings.

To conclude, the current study provides behavioral and neural evidence of interpersonal difficulties in boys with CD. The results suggest that CD (vs TD) boys do not make differential fairness decisions in response to different emotions of others, which is associated with reduced responses to others' emotions in brain regions important for social decision-making in CD (vs TD) boys.

Acknowledgements

We thank all participants and their parents. We also thank the participating centers (Palmhuis de Jutters, Forensisch Centrum Teylingereind) and Romy Emmerig and Simone van Montfort for their help with data collection.

Funding

This study was supported by the Netherlands Organization for Scientific Research (NWO) Grant No. 056-23-011 and by a Leiden Institute for Brain and Cognition-Starting Grant (O. Colins).

References

- Achenbach, T.M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: Department of Psychiatry, University of Vermont.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. 5th edn. Arlington, VA: American Psychiatric Association.
- Blair, R.J. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience*, *14*(11), 786–99.
- Carter, R.M., Huettel, S.A. (2013). A nexus model of the temporal-parietal junction. *Trends in Cognitive Sciences*, *17*(7), 328–36.
- Colins, O., Vermeiren, R., Vreugdenhil, C., van den Brink, W., Doreleijers, T., Broekaert, E. (2010). Psychiatric disorders in detained male adolescents: a systematic literature review. *Canadian Journal of Psychiatry*. *Revue Canadienne de Psychiatrie*, *55*(4), 255–63.
- Corbetta, M., Patel, G., Shulman, G.L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, *58*(3), 306–24.
- Dawel, A., O'Kearney, R., McKone, E., Palermo, R. (2012). Not just fear and sadness: meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and Biobehavioral Reviews*, *36*(10), 2288–304.
- de Castro, B.O., Veerman, J.W., Koops, W., Bosch, J.D., Monshouwer, H.J. (2002). Hostile attribution of intent and aggressive behavior: a meta-analysis. *Child Development*, *73*(3), 916–34.
- de Wied, M., van Boxtel, A., Matthys, W., Meeus, W. (2012). Verbal, facial and autonomic responses to empathy-eliciting film clips by disruptive male adolescents with high versus low callous-unemotional traits. *Journal of Abnormal Child Psychology*, *40*(2), 211–23.
- Dodge, K.A. (1993). Social-cognitive mechanisms in the development of conduct disorder and depression. *Annual Review of Psychology*, *44*(1), 559–84.
- Fairchild, G., van Goozen, S.H., Calder, A.J., Stollery, S.J., Goodyer, I.M. (2009). Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. *Journal of Child Psychology and Psychiatry*, *50*(5), 627–36.
- Falk, A., Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.
- Fanti, K.A., Panayiotou, G., Lombardo, M.V., Kyranides, M.N. (2016). Unemotional on all counts: evidence of reduced affective responses in individuals with high callous-unemotional traits across emotion systems and valences. *Social Neuroscience*, *11*, 72–87.
- Frick, P.J., Ray, J.V., Thornton, L.C., Kahn, R.E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, *140*(1), 1–57.
- Frith, C.D., Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–4.
- Gummerum, M., Hanoch, Y. (2012). Altruism behind bars: sharing, justice, perspective taking and empathy among inmates. *Social Justice Research*, *25*(1), 61–78.
- Güroğlu, B., van den Bos, W., Crone, E.A. (2009). Fairness considerations: increasing understanding of intentionality during adolescence. *Journal of Experimental Child Psychology*, *104*(4), 398–409.
- Happé, F., Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. *British Journal of Developmental Psychology*, *14*, 385–98.
- Hasler, G. (2012). Can the neuroeconomics revolution revolutionize psychiatry? *Neuroscience and Biobehavioral Reviews*, *36*(1), 64–78.
- Herpertz, S.C., Huebner, T., Marx, I., et al. (2008). Emotional processing in male adolescents with childhood-onset conduct disorder. *Journal of Child Psychology and Psychiatry*, *49*(7), 781–91.
- Herpertz, S.C., Mueller, B., Qunaibi, M., Lichtenfeld, C., Konrad, K., Herpertz-Dahlmann, B. (2005). Response to emotional stimuli in boys with conduct disorder. *American Journal of Psychiatry*, *162*(6), 1100–7.
- Johnston, L., Hawes, D.J., Straiton, M. (2014). Psychopathic traits and social cooperation in the context of emotional feedback. *Psychiatry, Psychology and Law*, *21*, 767–78.
- Jolliffe, D., Farrington, D.P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589–611.
- Jones, A.P., Laurens, K.R., Herba, C.M., Barker, G.J., Viding, E. (2009). Amygdala hypoactivity to fearful faces in boys with conduct problems and callous-unemotional traits. *American Journal of Psychiatry*, *166*(1), 95–102.
- Kahneman, D., Knetsch, J.L., Thaler, R.H. (1986). Fairness and the assumptions of economics. *Journal of Business*, *59*(4), S285–300.
- Kaufman, J., Birmaher, B., Brent, D., et al. (1997). Schedule for affective disorders and schizophrenia for school-age children present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(7), 980–8.
- Kimonis, E.R., Frick, P.J., Skeem, J.L., et al. (2008). Assessing callous-unemotional traits in adolescent offenders: validation of the Inventory of Callous-Unemotional Traits. *International Journal of Law and Psychiatry*, *31*(3), 241–52.

- Kishida, K.T., King-Casas, B., Montague, P.R. (2010). Neuroeconomic approaches to mental disorders. *Neuron*, *67*(4), 543–54.
- Klapwijk, E.T., Peters, S., Vermeiren, R.R., Lelieveld, G.J. (2013). Emotional reactions of peers influence decisions about fairness in adolescence. *Frontiers in Human Neuroscience*, *7*, 745.
- Koenigs, M., Kruepke, M., Newman, J.P. (2010). Economic decision-making in psychopathy: a comparison with ventromedial prefrontal lesion patients. *Neuropsychologia*, *48*(7), 2198–204.
- Krall, S.C., Rottschy, C., Oberwelland, E., et al. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*(2), 587–604.
- Lelieveld, G.J., van Dijk, E., Guroglu, B., et al. (2013a). Behavioral and neural reactions to emotions of others in the distribution of resources. *Social Neuroscience*, *8*(1), 52–62.
- Lelieveld, G.J., van Dijk, E., van Beest, I., van Kleef, G.A. (2012). Why anger and disappointment affect other's bargaining behavior differently: the moderating role of power and the mediating role of reciprocal and complementary emotions. *Personality and Social Psychology Bulletin*, *38*(9), 1209–21.
- Lelieveld, G.J., van Dijk, E., van Beest, I., van Kleef, G.A. (2013b). Does communicating disappointment in negotiations help or hurt? Solving an apparent inconsistency in the social-functional approach to emotions. *Journal of Personality and Social Psychology*, *105*(4), 605–20.
- Lieberman, M.D., Cunningham, W.A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*(4), 423–8.
- Lockwood, P.L., Sebastian, C.L., McCrory, E.J., et al. (2013). Association of callous traits with reduced neural response to others' pain in children with conduct problems. *Current Biology*, *23*(10), 901–5.
- Marsh, A.A., Finger, E.C., Fowler, K.A., et al. (2013). Empathic responsiveness in amygdala and anterior cingulate cortex in youths with psychopathic traits. *Journal of Child Psychology and Psychiatry*, *54*(8), 900–10.
- Miller, E.K., Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202.
- Mitchell, J.P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*(2), 262–71.
- Mokros, A., Menner, B., Eisenbarth, H., Alpers, G.W., Lange, K.W., Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner's dilemma game yields higher rewards. *Journal of Abnormal Psychology*, *117*(2), 406–13.
- Ochsner, K.N., Silvers, J.A., Buhle, J.T. (2012). Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, *1251*, E1–24.
- Pardini, D. (2011). Perceptions of social conflicts among incarcerated adolescents with callous-unemotional traits: 'you're going to pay. It's going to hurt, but I don't care.'. *Journal of Child Psychology and Psychiatry*, *52*(3), 248–55.
- Radke, S., Brazil, I.A., Scheper, I., Bulten, B.H., de Bruijn, E.R. (2013). Unfair offers, unfair offenders? Fairness considerations in incarcerated individuals with and without psychopathy. *Frontiers in Human Neuroscience*, *7*, 406.
- Rilling, J.K., Glenn, A.L., Jairam, M.R., et al. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological Psychiatry*, *61*(11), 1260–71.
- Rilling, J.K., Sanfey, A.G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, *62*(1), 23–48.
- Schwenck, C., Mergenthaler, J., Keller, K., et al. (2012). Empathy in children with autism and conduct disorder: group-specific profiles and developmental aspects. *Journal of Child Psychology and Psychiatry*, *53*(6), 651–9.
- Sebastian, C.L., McCrory, E.J.P., Cecil, C.A.M., et al. (2012). Neural responses to affective and cognitive theory of mind in children with conduct problems and varying levels of callous-unemotional traits. *Archives of General Psychiatry*, *69*(8), 814–22.
- Sharp, C., Burton, P.C., Ha, C. (2011). "Better the devil you know": a preliminary study of the differential modulating effects of reputation on reward processing for boys with and without externalizing behavior problems. *European Child and Adolescent Psychiatry*, *20*(11-12), 581–92.
- Sharp, C., Ha, C., Fonagy, P. (2011). Get them before they get you: trust, trustworthiness, and social cognition in boys with and without externalizing behavior problems. *Development and Psychopathology*, *23*(2), 647–58.
- Silani, G., Lamm, C., Ruff, C.C., Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *Journal of Neuroscience*, *33*(39), 15466–76.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, *56*(1), 185–96.
- Steinbeis, N., Bernhardt, B.C., Singer, T. (2012). Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron*, *73*(5), 1040–51.
- van den Bos, W., Vahl, P., Guroglu, B., et al. (2014). Neural correlates of social decision-making in severely antisocial adolescents. *Social Cognitive and Affective Neuroscience*, *9*(12), 2059–66.
- van Kleef, G.A., De Dreu, C.K.W., Manstead, A.S.R. (2010). An interpersonal approach to emotion in social decision making. *Advances in Experimental Social Psychology*, *42*, 45–96.
- Van Overwalle, F., Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564–84.
- White, S.F., Fowler, K.A., Sinclair, S., et al. (2014). Disrupted expected value signaling in youth with disruptive behavior disorders to environmental reinforcers. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(5), 579–88.e579.
- Woo, C.W., Krishnan, A., Wager, T.D. (2014). Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. *Neuroimage*, *91*, 412–9.