

Toddler–mother attachment moderates adolescents' behavioral and neural evaluation of trustworthiness

Xiaomei Li¹ Nathan A. Jorgensen,² Nancy L. McElwain,^{1,3} and Eva H. Telzer²

¹Department of Human Development and Family Studies, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

²Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270, USA

³Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Correspondence should be addressed to Xiaomei Li, Department of Human Development and Family Studies, University of Illinois at Urbana-Champaign, 2027 Christopher Hall, MC-018, 904 West Nevada Street, Urbana, IL 61801, USA. E-mail: xiaomei3@illinois.edu.

Abstract

This longitudinal study examined the prospective association between toddler–mother attachment to adolescents' ($n = 52$; 34 boys; $M_{\text{age}} = 13.22$ years; 90% White) behavioral and neural responses during the evaluation of trustworthiness from unfamiliar, emotionally neutral faces. At 33 months, toddler–mother attachment status (secure vs insecure classification) was assessed using a modified Strange Situation procedure. Results revealed that attachment moderated the processing of trustworthiness facial cues. As faces became less trustworthy, adolescents with a secure (vs insecure) attachment history rated the faces as correspondingly less trustworthy and showed increasing (vs overall blunted) activation in brain regions involved in trustworthiness perception (i.e. bilateral amygdala, bilateral fusiform, right anterior insula and right posterior superior temporal sulcus). Findings suggest that a secure compared with insecure child–mother attachment in toddlerhood may be associated with greater capacity for, or openness to, processing potentially negative social information at both the behavioral and neural levels during adolescence.

Key words: attachment; trustworthiness; social information processing; developmental social neuroscience; adolescence

Interpersonal trust, considered as a generalized expectancy that others can be relied on (Rotter, 1971), is central to healthy social relationships across the lifespan (Simpson, 2007). Among youth, holding an accurate level of trust in others has been linked with better social competence and adjustment, including more prosocial behavior, higher peer acceptance, less social exclusion and fewer internalizing symptoms (Wentzel, 1991; Rotenberg *et al.*, 2004, 2005; Carlo *et al.*, 2010). Yet, individual differences in adolescents' ability to evaluate trustworthiness of unfamiliar faces—an attributional process that occurs rapidly over milliseconds (Willis and Todorov, 2006; Todorov *et al.*, 2009)—have been observed at both behavioral (i.e. the extent to which one rates others as trustworthy) and neural (i.e. variability in brain activity when evaluating trustworthiness of faces; e.g. Nowakowski *et al.*, 2010) levels of analysis. Informed by prior work linking the quality of early attachment relationships and subsequent social information processing (SIP) patterns (e.g. Dykas and Cassidy, 2011; Vrtička and Vuilleumier, 2012; Zimmermann and Iwanski, 2015; Long *et al.*, 2020), the current study aimed to examine prospective links between toddler–mother attachment and adolescents' evaluation of trustworthiness.

Several key brain regions have been proposed to underly evaluation of trustworthiness in emotionally neutral faces (see Bzdok

et al., 2011 for a meta-analysis), including (a) the amygdala, given its role in processing information of motivational relevance (Adolphs *et al.*, 2003); (b) the anterior insula (AI), given its role in salience detection through the mapping of autonomic changes in the body (Menon and Uddin, 2010); (c) the fusiform gyrus, given its involvement in face processing and social perception (Kanwisher *et al.*, 1997; Haxby *et al.*, 2002); and (d) the posterior superior temporal sulcus (pSTS), given its involvement in mentalizing (Frith and Frith, 2003; Saxe *et al.*, 2004). Indeed, among adults, when viewing emotionally neutral faces that differed in the level of trustworthiness, studies consistently revealed more activation in several key brain regions (i.e. bilateral amygdala, right AI, and right pSTS) in response to untrustworthy compared with trustworthy faces (e.g. Winston *et al.*, 2002; Engell *et al.*, 2007; Mende-Siedlecki *et al.*, 2013), suggesting that brain regions broadly involved in motivational processing, salience detection and mentalizing tend to (a) track variations in trustworthiness from unfamiliar faces and (b) show enhanced responsivity to increasingly untrustworthy faces, perhaps due to potential signals of social threat evoked by such faces. Only one study, to date, has examined neural correlates of trustworthiness perception among 'adolescents'. Similar to adults, adolescents demonstrated increased activation to untrustworthy, emotionally neutral faces

Received: 5 April 2021; Revised: 23 November 2021; Accepted: 31 January 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

in the bilateral amygdala, bilateral fusiform gyrus and right AI (Kragel et al., 2015), revealing that adolescents also associate untrustworthy faces with heightened motivational relevance and social salience.

To understand factors related to individual differences in adolescent trustworthiness evaluation, we were guided by attachment theory in which a core component of child-caregiver attachment is trust in the caregiver's availability (Bowlby, 1973, 1980; Sroufe and Fleeson, 1986). Early attachment relationships with caregivers have been proposed to shape differential patterns of SIP more broadly, both at behavioral (Dykas and Cassidy, 2011; Zimmermann and Iwanski, 2015) and at neural (Vrtička and Vuilleumier, 2012; Long et al., 2020) levels. Dykas and Cassidy (2011; also see Zimmermann & Iwanski, 2015) proposed that individuals classified as secure process a broad range of positive and negative social information in an open manner due to their greater capacity to explore the environment and tolerate aversive situations, whereas individuals classified as insecure tend to engage in less open and flexible evaluation of negative social information in particular. Eye-tracking studies found that youth who reported a secure vs insecure attachment style had more fixations on, and longer viewing times of, their mothers' faces characterized by both positive and negative emotional valence (Vandevivere et al., 2014). Conversely, insecure vs secure attachment was associated with shorter looking time at neutral and emotionally negative facial expressions among children (Kammermeier et al., 2019).

Intriguingly, two studies provide neural evidence in line with the above behavioral studies. Adolescent-reported attachment avoidance was associated with decreased activation to emotionally conflicting social feedback (e.g. angry face displayed next to a message saying 'winning') in the left amygdala and right AI (Vrtička et al., 2014). Similar to Dykas and Cassidy's (2011) propositions, the authors concluded that low activation among more avoidant adolescents may reflect decreased social sensitivity in general, or alternatively, a protective strategy to attribute less self-relevance and affective salience to emotionally negative social information. Similarly, Escobar et al. (2013) reported that insecure vs secure adolescents showed less accuracy and more difficulties, as reflected in event-related potentials, differentiating facial expressions characterized by negative emotional valence. Nonetheless, despite emerging neural evidence of (i) adolescents' perception of trustworthiness cues and (ii) attachment-related differences in SIP, no prior study has linked the two. Moreover, despite theoretical (e.g. Bowlby, 1973, 1980; Dykas and Cassidy, 2011) and empirical arguments (e.g. Corriveau et al., 2009) for examining how 'early' attachment relationships shape 'later' SIP patterns (in our case, trustworthiness perception) from a neural perspective, prior studies on this topic are largely cross-sectional and utilize self-report measures of attachment (e.g. Escobar et al., 2013; Vrtička et al., 2014).

Using data from a 10 year longitudinal study, we aimed to address these gaps and examined whether toddler-mother attachment relationships assessed from a modified Strange Situation procedure was associated with perception of trustworthiness from unfamiliar faces during early adolescence. Informed by Dykas and Cassidy's (2011) propositions, we posit that individual differences in trustworthiness perception are rooted in adolescents' experience-based, attachment-related internal working models. A history of warm and reliable caregiving likely establishes a basis for mental representation of others as approachable in novel and ambiguous interpersonal contexts, as well as bolsters confidence in exploration (in this case, cognitive evaluation) in the

context of both socially positive/pleasant and negative/aversive cues (e.g. more and less trustworthy faces, respectively; Schore, 2001; Bretherton and Munholland, 2008). Conversely, a history of inconsistent or rejecting caregiving may give rise to expectations that others are not reliable and, in turn, hinder processing of aversive social information, such as less trustworthy faces (Schore, 2001; Bretherton and Munholland, 2008).

Furthermore, we assessed adolescents' behavioral and neural trustworthiness perception during a functional magnetic resonance imaging (fMRI) task that involved viewing facial stimuli that represented a range of standardized, externally validated trustworthiness ratings (Lundqvist et al., 1998). Guided by prior neural investigations of trustworthiness perception (Winston et al., 2002; Engell et al., 2007; Bzdok et al., 2011; Mende-Siedlecki et al., 2013; Kragel et al., 2015), we adopted an a priori regions of interest (ROI) approach to assess trustworthiness perception in response to emotionally neutral faces, in which we expected attachment-group differences in neural tracking of trustworthiness in the bilateral amygdala, bilateral fusiform gyrus, right AI and right pSTS. We hypothesized that secure (vs insecure) attachment would be associated with greater decreases in adolescents' trustworthiness ratings as the trustworthiness of faces decreased. Likewise, we expected secure (vs insecure) attachment to be linked with adolescents' increased neural activation as faces became less trustworthy.

Method

Participants

Data were drawn from a 10 year longitudinal study of socioemotional development named 'Children's Social Development Project' (CSDP). All research protocols pertaining to the CSDP were reviewed and approved by the Institutional Review Board at the University of Illinois at Urbana-Champaign (protocols #05181, #15435). Families were recruited via informational flyers and birth announcements distributed through local organizations and child care centers. At the initial time point, 128 toddlers (62 boys; M age = 32.7 months, $s.d.$ = 0.76; 52% first born) and their mothers participated in a 90 minute laboratory visit (see McElwain et al., 2012 for further details). When children were approximately 13 years of age, families were contacted to participate in a follow-up study. Adolescents who returned ($n = 67$) vs declined to participate ($n = 61$) were more likely to be boys, $\chi^2(1) = 11.43$, $P = 0.001$. Fathers of adolescents who returned tended to have fewer years of education ($M = 15.5$ years, $s.d.$ = 2.36) compared with those who did not ($M = 16.6$ years, $s.d.$ = 2.90), $t(109.07) = 2.373$, $P = 0.019$. No differences between the two groups were found for toddler-mother attachment, parental age and ethnicity, maternal education or marital status.

Of the 67 families who participated in the adolescent phase, which included a behavioral session (see Ravindran et al., 2020 for further details) and a neuroimaging session, 52 adolescents (34 boys; M age = 13.2 years, $s.d.$ = 0.57, range = 12.4–14.8 years) successfully completed the neuroimaging session. Reasons for incompleteness included MRI contraindications (braces, $n = 7$; claustrophobia, $n = 2$) and declining to participate ($n = 6$). Approximately 90% of the adolescent sample ($n = 52$) were identified by mothers as European American, 8% as African American and 2% as mixed or more than one ethnicity. Mothers were mostly biological (96%) and married (94%) at the initial time point. Mothers and fathers averaged 33.1 ($s.d.$ = 5.90) and 34.4 ($s.d.$ = 5.60) years of age, respectively, and had 16.3 ($s.d.$ = 1.86) and 15.6 ($s.d.$ = 2.26) years of education. Mothers and fathers were 81%

and 89% European American, 2% and 4% African American, 6% and 0% Asian American, 2% and 0% Hispanic, and 8% and 2% others such as Native American or biracial, respectively. Adolescents who completed ($n=52$) vs did not complete ($n=15$) the neuroimaging session did not differ on toddler–mother attachment, age at the adolescent phase, biological sex or parental age, ethnicity, education or marital status.

Measures

Toddler–mother attachment (33 months)

Cassidy *et al.* (1992) modified 17 min Strange Situation procedure, which consists of five episodes (3 min warm-up, 3 min separation, 3 min reunion, second 5 min separation, second 3 min reunion), was conducted at the beginning of the laboratory visit at the initial time point. During the separation episodes, no ‘stranger’ was present, and mothers received no instructions about what to tell their child during the departure from the playroom. Toddler behavior was evaluated using established criteria, including physical proximity to the mother, affective expression and verbal exchanges (Cassidy *et al.*, 1992). The Cassidy *et al.* (1992) system has established validity and is considered the measure of choice for assessing attachment security among children between 2.5 and 4.5 years of age (see Solomon and George, 2008).

Two trained coders, certified by Jude Cassidy, coded protocols from the full sample of 128 toddlers. To assess interobserver reliability, 20% of the protocols were double-coded, and disagreements were resolved by consensus. Interobserver agreement (before consensus) was 88% ($\kappa = 0.77$) for the four-way classification and 92% ($\kappa = 0.83$) for the two-way secure vs insecure classification. For the current subsample of 52, toddlers were classified as ‘secure’ ($n=33$), ‘insecure-avoidant’ ($n=2$), ‘insecure-ambivalent or dependent’ ($n=8$), or ‘controlling or insecure other’ ($n=9$; see NICHD Early Child Care Research Network, 2001, for similar proportions of secure and insecure classifications using the same procedures at 36 months). Given the small cell sizes of the insecure classifications, we used the binary secure–insecure classification.

Evaluation of trustworthiness (13 years)

While undergoing an fMRI scan, adolescents’ behavioral and neural responses to trustworthiness faces were assessed. Facial stimuli were drawn from the Karolinska Directed Emotional Faces dataset and represent a range of standardized, externally validated trustworthiness ratings (Lundqvist *et al.*, 1998), which we refer to as ‘consensus ratings’ (same term as in Engell *et al.*, 2007). We selected 28 faces (14 female) to capture a range of trustworthiness consensus ratings from untrustworthy to trustworthy (range: -1.51 to 1.35 ; $M=0.024$; no difference between male and female faces). Adolescents saw four randomized blocks of the 28 faces, with each block corresponding to a particular emotion (i.e. neutral, happy, disgusted and sad). Consistent with prior studies that used comparable paradigms and stimuli to investigate trustworthiness perception (Winston *et al.*, 2002; Engell *et al.*, 2007; Mende-Siedlecki *et al.*, 2013; Kragel *et al.*, 2015), we examined neural and behavioral responses from the block that contained emotionally neutral faces only. Within blocks, trial order was pre-randomized such that each participant saw the same sequence of faces within each block, and adolescents rated each of the 28 faces on trustworthiness using a 5-point scale (1 = ‘very untrustworthy’ to 5 = ‘very trustworthy’). Adolescents were told to think of trustworthiness in terms of being alone in an unfamiliar city

and being willing to approach the person to ask for help or directions. They used their right hand for all responses with the thumb indicating ‘1’ and the pinky finger indicating ‘5.’ Each face was presented for 2500 ms with a jittered crosshair (centered with a gamma distribution around 1500 ms).

fMRI data acquisition. Imaging data were collected using a 3 Tesla Siemens Magnetom Trio MRI scanner. The trust task was presented on a computer screen and projected through a mirror. A high-resolution structural T2*-weighted echoplanar imaging (EPI) volume (TR = 2000 ms; TE = 25 ms; matrix = 92×92 ; FOV = 230 mm; 38 slices; slice thickness = 3 mm; voxel size $2.5 \times 2.5 \times 3$ mm³) was acquired coplanar with a T2*-weighted structural matched-bandwidth (MBW), high-resolution, anatomical scan (TR = 4000 ms; TE = 64 ms; matrix = 192×192 ; FOV = 230 mm; voxel size $1.2 \times 1.2 \times 3$ mm³; 38 slices; slice thickness = 3 mm). In addition, a T1* magnetization-prepared rapid-acquisition gradient echo (MPRAGE; TR = 1900 ms; TE = 2.32 ms; matrix = 256×256 ; FOV = 230 mm; voxel size $0.9 \times 0.9 \times 0.9$ mm³; sagittal plane; slice thickness = 0.9 mm; 192 slices) was acquired. The orientation for the EPI and MBW scans was oblique axial to maximize brain coverage and to reduce noise.

fMRI data analyses. Preprocessing used FSL (FMRIB’s Software Library, version 6.0; www.fmrib.ox.ac.uk/fsl) and included the following steps: skull stripping using BET (Smith, 2002); motion correction with MCFLIRT (Jenkinson *et al.*, 2002); spatial smoothing with Gaussian kernel of full width at half maximum 6 mm; high-pass temporal filtering with a filter width of 128 s (Gaussian-weighted least-squares straight line fitting, with $\sigma = 64.0$ s); grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; and individual level ICA denoising for motion and physiological noise using MELODIC (version 3.15; Beckmann and Smith, 2004), combined with an automated signal classifier (Tohka *et al.*, 2008; Neyman–Pearson threshold = 0.3). Functional images were resampled to a $2 \times 2 \times 2$ mm space and coregistered in a two-step sequence to the MBW and the MPRAGE images using FLIRT to warp them into the standard stereotaxic space defined by the Montreal Neurological Institute and the International Consortium for Brain Mapping.

Individual-level, fixed-effects analyses were estimated using the general linear model convolved with a canonical hemodynamic response function in SPM8. The task was modeled as event-related within the emotionally neutral block, with each trial lasting for the duration of the image. The consensus rating for each image was used as a parametric modulator (PM) for each trial, allowing us to examine neural regions that show linear increases or decreases in sensitivity to the trustworthiness ratings. The jittered inter-trial periods were not modeled and served as the implicit baseline for the task. Six motion parameters were modeled as regressors of no interest.

Our primary, confirmatory analyses employed an ROI approach with four masks (see Figure 1): bilateral amygdala (Harvard–Oxford atlas thresholded at 50%), bilateral fusiform gyrus (combined OFA and FFA; Julian *et al.*, 2012), right AI (Harvard–Oxford atlas) and right pSTS (Julian *et al.*, 2012). Masks are available on Neurovault (Gorgolewski *et al.*, 2015; <https://neurovault.org/collections/IRRADID/>). Using these masks, we extracted parameter estimates of signal intensity from the neutral facial stimuli for each brain region. To represent our primary construct of interest—neural tracking of trustworthiness—parameter estimates were extracted from neutral faces including the PM, which indicates neural activation

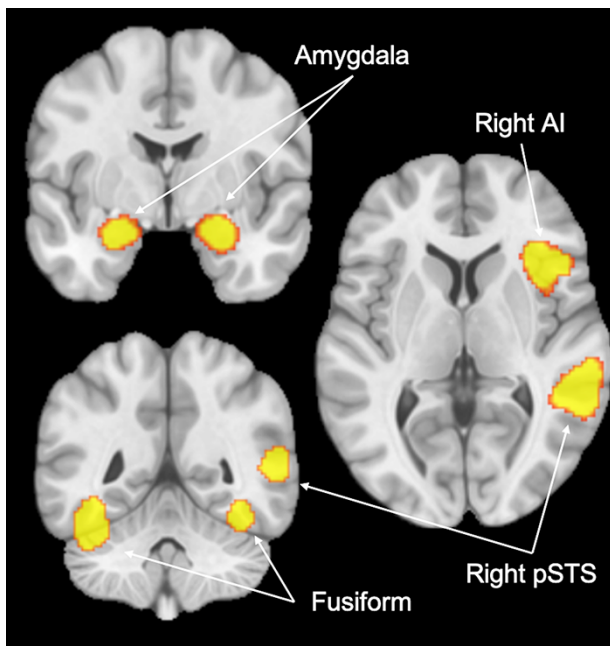


Fig. 1. Regions of Interest (ROI) images for the amygdala, right anterior insula (AI), right posterior superior temporal sulcus (pSTS) and fusiform gyrus.

Note: Although presented together in this image, each ROI was extracted separately in analyses.

associated with changes in the consensus trustworthiness ratings (termed ‘modulated activation’ in this report). To control for adolescents’ average brain activation across faces, parameter estimates were extracted from the neutral faces controlling for the PM, which represents neural activation for the average consensus rating of trustworthiness (i.e. $M=0$; termed ‘baseline activation’ in this report).

Data analytic plan

To test our hypotheses, we fit a series of models in Mplus 8.1 (Muthén and Muthén, 2017) using robust maximum likelihood estimation. Prior to the main model tests, we conducted preliminary analyses to examine whether key study variables differ as a function of adolescent age and biological sex given their implication in adolescent processing of social information (e.g. Vrtička et al., 2014), as well as maternal years of education as a proxy measure of family socioeconomic status. Demographic variables that shared a significant association with key outcomes (adolescents’ ratings and modulated activation) were included as covariates in the main model tests.

To assess attachment-group differences in adolescents’ ratings of trustworthiness, we fitted a series of multilevel models, with trials nested within persons. We first tested an intercept-only model, which indicated sufficient variability in adolescents’ trustworthiness ratings at both the within-person (0.917) and between-person (0.292) levels; the intraclass correlation coefficient (ICC) was 24.15%, indicating that 24% and 76% of the variation in trustworthiness ratings was between and within persons, respectively. Thus, we proceeded with the test of the conditional model. At level 1, we modeled the consensus ratings as a predictor of adolescents’ trustworthiness ratings across trials. This within-person association indicated the degree to

which adolescents’ ratings of trustworthiness tracked the consensus ratings. At level 2, we tested attachment group (insecure = 0, secure = 1) as a predictor of the intercept of adolescents’ ratings (i.e. average rating of trustworthiness across trials) and the level-1 random slope. The cross-level interaction between attachment at level 2 and the slope at level 1 assessed whether trustworthiness evaluations (as indexed by the association between the consensus and adolescents’ ratings) differed as a function of toddler–mother attachment. The consensus ratings were mean-centered prior to model tests. The covariance between the intercept and level-1 random slope was also estimated at level 2.

To assess attachment-group differences in adolescents’ modulated neural activation (i.e. neural activation associated with changes in consensus ratings), we fitted a series of multilevel models, with ROIs nested within persons. We integrated tests of four ROIs simultaneously in one multilevel model given prior work indicating similar patterns of modulated activation (i.e. increased activation to trustworthy vs untrustworthy faces) and to minimize multiple comparison biases caused by fitting a separate model for each ROI. The intercept-only model indicated significant variability in adolescents’ overall neural tracking of trustworthiness at both the within-person (0.043) and between-person (0.043) levels; ICC = 50%. For the conditional model, at level 1, we modeled baseline activation and a set of ROI binary variables as predictors of the modulated activation. The four ROIs were recoded into three binary variables using effect coding, in which the fusiform, amygdala and AI were each contrasted with the pSTS. For each ROI binary variable, the selected region was coded as 1, the contrast region (i.e. pSTS) as -1, and the remaining two regions as 0. When the three ROI binary variables are entered together in the model, coefficients for other predictors represent the parameter estimates across all ROIs. At level 2, we tested baseline activation and attachment group (insecure = 0, secure = 1) as predictors of the intercept of the modulated activation. The level-2 association between attachment and modulated activation indicated whether the attachment groups differ on neural tracking of trustworthiness. In a preliminary multilevel model, we also tested attachment group as a level-2 predictor of a random slope between the ROI binary variables and the modulated activation at level 1. This cross-level interaction, which tested whether the association between attachment and modulated activation differed as a function of ROI, was non-significant ($B = -0.01$, $SE = 0.02$, $P = 0.610$) and thus was excluded in the main multilevel model reported below.

Results

Preliminary analyses

Data were checked for distribution, outliers, and missingness in SPSS 25.0. Four percent ($n = 59$ out of 1456 data points) of adolescents’ ratings of trustworthiness were missing because the adolescent did not respond within the trial time window. Little’s (1988) MCAR test suggested that data were likely missing completely at random, $\chi^2(3) = 0.76$, $P = 0.860$, and thus were handled using full information maximum likelihood in the main model tests. Descriptive statistics and bivariate correlations among the continuous study variables are summarized in Table 1.

To assess whether demographic variables should be included in main model tests as potential covariates, we examined whether adolescents’ trustworthiness ratings and modulated activation differed as a function of adolescent age and biological sex and

Table 1. Descriptive statistics and bivariate correlations among continuous study variables

Study variables	1	2	3	4	5	6	7	8	9
1 Adolescents' ratings ^a	—								
Baseline activation									
2 Bilateral amygdala	-0.08	—							
3 Bilateral fusiform	-0.11	0.65***	—						
4 Right AI	0.07	0.23	0.23 [†]	—					
5 Right pSTS	-0.12	0.50***	0.36**	0.67***	—				
Modulated activation									
6 Bilateral amygdala	0.04	-0.01	-0.11	0.05	0.09	—			
7 Bilateral fusiform	-0.03	0.03	-0.11	0.11	0.08	0.77***	—		
8 Right AI	0.01	-0.11	0.01	0.06	-0.07	0.61***	0.62***	—	
9 Right pSTS	-0.13	0.05	0.09	-0.21	-0.13	0.25 [†]	0.33*	0.52***	—
M	2.67	0.42	0.52	0.01	-0.06	-0.04	-0.02	0.04	0.05
s.d.	0.58	0.72	0.60	0.65	0.56	0.32	0.28	0.29	0.28
Min.	1.46	-1.14	-0.88	-1.57	-1.39	-0.86	-0.66	-0.53	-0.53
Max.	4.17	2.58	2.02	1.71	1.22	0.98	0.76	0.97	0.82

Notes: AI = anterior insula; pSTS = posterior superior temporal sulcus. Baseline activation represents the average neural activation in each brain region across all faces. Modulated activation represents neural activation associated with increases in the consensus trustworthiness rating for each brain region.

^aAdolescents' rating represented the person 'mean' of trustworthiness ratings across trials for each adolescent.
[†] $P < 0.10$. * $P < 0.05$. ** $P < 0.01$. *** $P < 0.001$, two-tailed.

maternal education. No significant difference was found. Therefore, no demographic variable was included in the main model tests.

We also examined attachment-group differences in baseline activation. Independent samples *t*-tests indicated one attachment-group difference: Baseline activation of the bilateral fusiform was higher among adolescents with a secure ($M = 0.66$, $s.d. = 0.52$) vs insecure ($M = 0.27$, $s.d. = 0.67$) attachment history, $t(30.81) = -2.17$, $P = 0.038$.

Main model tests

Behavioral ratings of trustworthiness

Results from the multilevel model predicting adolescents' trustworthiness ratings are shown in Table 2. At level 1, the random linear slope was significant such that higher consensus ratings were associated with higher adolescents' ratings, indicating that adolescents were perceiving more trustworthy faces as more trustworthy accordingly. The cross-level interaction between attachment at level 2 and consensus ratings at level 1 was significant, indicating that the strength of the level-1 association between the consensus and adolescent ratings differed as a function of attachment.

To probe this significant cross-level interaction, we examined the simple slope for each attachment group (i.e. the level-1 association when attachment = 0). Because the insecure group was coded as 0 in the main model, the level-1 slope reported in Table 2 represents the simple slope for the insecure group. To obtain the simple slope for the secure group, we fit an identical multilevel model with attachment reverse-coded (secure = 0, insecure = 1). As shown in Figure 2, these simple slopes revealed that although adolescents in both attachment groups showed a positive association between the consensus and adolescents' ratings of trustworthiness, the association was weaker for adolescents in the insecure ($B = 0.32$, $SE = 0.07$, $P < 0.001$) compared with secure group ($B = 0.50$, $SE = 0.05$, $P < 0.001$). To further probe the interaction pattern, we examined attachment-group differences in adolescents' ratings at lower (1 s.d. below mean; less trustworthy) and higher (1 s.d. above mean; more trustworthy) levels of the consensus ratings. Adolescents in the insecure (vs secure) group rated less trustworthy faces as significantly higher on trustworthiness ($B = -0.37$, $SE = 0.19$, $P = 0.047$), but the two groups did not

Table 2. Adolescents' ratings of trustworthiness as a function of the consensus ratings and toddler-mother attachment groups

	B	SE	P
Intercept (adolescent ratings)	2.81	0.14	<0.001
Level 1			
Consensus ratings → adolescent ratings (slope)	0.32	0.07	<0.001
Level 2 predictors of intercept			
Attachment (insecure = 0, secure = 1)	-0.23	0.17	0.162
Level 2 predictors × level 1 slope cross-level interactions			
Attachment × consensus ratings	0.18	0.09	0.037
Covariance (level 1 slope with intercept)	-0.02	0.02	0.259
Random effects			
Residual variance (level 1): adolescent ratings	0.77	0.10	<0.001
Residual variance (level 2): intercept	0.28	0.06	<0.001
Residual variance (level 2): Level-1 slope	0.05	0.02	0.004

differ in ratings of more trustworthy faces ($B = -0.09$, $SE = 0.17$, $P = 0.586$).

Neural processing of trustworthiness

Results from the multilevel model predicting adolescents' neural tracking of trustworthiness are shown in Table 3. The cross-level interaction between attachment (level 2) and the intercept of modulated activation (level 1) was significant, indicating that adolescents' neural tracking of trustworthiness differed as a function of attachment group. Baseline activation (levels 1 and 2) and the ROI variables (level 1) were not associated with modulated activation, indicating that neural tracking of trustworthiness was not related to the average activation across faces and did not differ as a function of region.

To probe the significant cross-level interaction, we fitted additional multilevel models with the ROI binary variables and attachment group at level 1 and level 2, respectively, predicting (i)

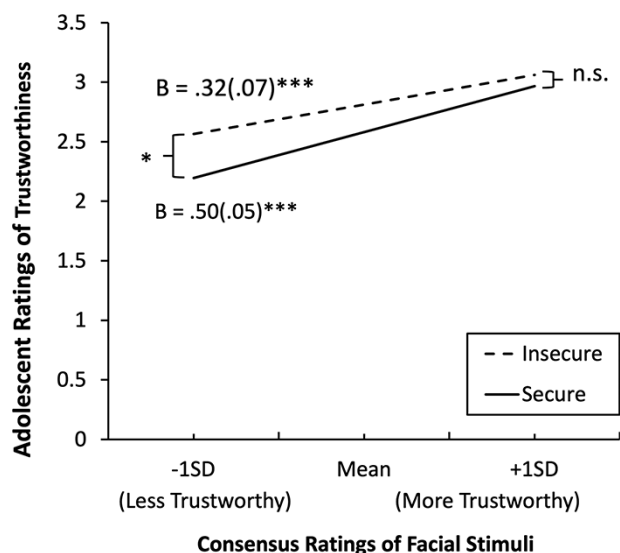


Fig. 2. Adolescents' ratings of trustworthiness as a function of the consensus ratings and toddler–mother attachment groups.

Notes: Brackets represent tests of attachment-group differences in adolescents' ratings at lower (1 s.d. below mean; less trustworthy) and higher (1 s.d. above mean; more trustworthy) levels of the consensus ratings. * $P < 0.05$. *** $P < 0.001$, two-tailed.

Table 3. Adolescents' neural tracking of trustworthiness from unfamiliar faces as a function of the consensus ratings and toddler–mother attachment groups

	B	SE	P
Intercept (modulated activation)	0.12	0.07	0.058
Level 1			
Baseline activation (level 1) → modulated activation	-0.03	0.04	0.496
Bilateral fusiform—right pSTS → modulated activation	-0.02	0.02	0.445
Bilateral amygdala—right pSTS → modulated activation	-0.05	0.02	0.062
Right AI—right pSTS → modulated activation	0.03	0.03	0.287
Level 2 predictors of intercept			
Baseline activation (level 2)	0.06	0.09	0.521
Attachment (insecure = 0, secure = 1)	-0.20	0.08	0.009
Random effects			
Residual variance (level 1): modulated activation	0.04	0.01	<0.001
Residual variance (level 2): intercept	0.03	0.01	0.007

Notes: AI = anterior insula; pSTS = posterior superior temporal sulcus. Baseline activation represents the average neural activation in each brain region across all faces. Modulated activation represents neural activation associated with increases in the consensus trustworthiness rating for each brain region.

baseline activation to obtain the value of neural activation when trustworthiness = 0 (i.e. mean) and (ii) modulated activation to obtain the value for the association between consensus ratings and neural activation for each attachment group (i.e. the simple slope). When attachment group was coded as insecure = 0 and secure = 1, the intercepts in these two models were used to plot the simple slope for the insecure group. We fitted an identical set of models with attachment group reverse-coded (secure = 0, insecure = 1) to obtain the plotting parameters for

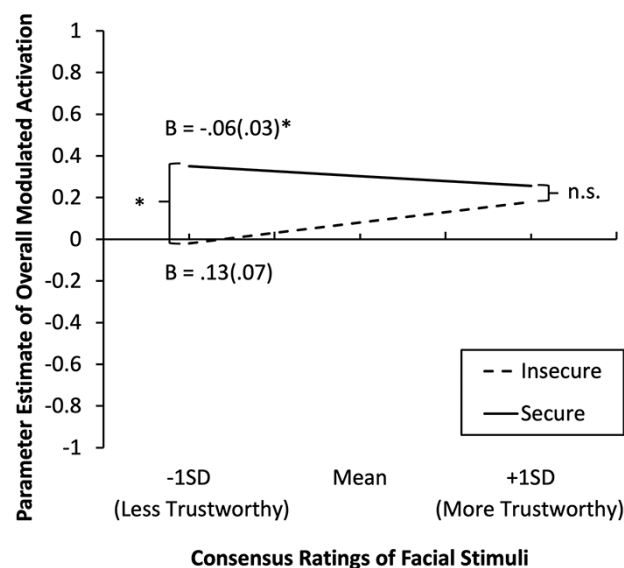


Fig. 3. Adolescent neural activation to trustworthiness as a function of the consensus ratings and toddler–mother attachment groups.

Notes: Associations were obtained from overall multilevel models fitted across all four regions of interest (ROI). Attachment-related differences in adolescents' neural tracking of trustworthiness did not differ across ROIs. Brackets represent tests of attachment-group differences in adolescents' overall modulated activation at lower (1 s.d. below mean; less trustworthy) and higher (1 s.d. above mean; more trustworthy) levels of the consensus ratings. * $P < 0.05$, two-tailed.

the secure group. As shown in Figure 3, adolescents with a secure attachment history showed a decrease in neural activation in response to faces as consensus ratings of trustworthiness increased ($B = -0.06$, $SE = 0.03$, $P = 0.026$), whereas this association was non-significant for adolescents with an insecure attachment history ($B = 0.13$, $SE = 0.07$, $P = 0.050$). To further probe the interaction pattern, we examined attachment-group differences in modulated activation at lower (1 s.d. below mean; less trustworthy) and higher (1 s.d. above mean; more trustworthy) levels of the consensus ratings. Adolescents in the insecure (vs secure) group showed greater neural activation to less trustworthy faces ($B = 0.37$, $SE = 0.14$, $P = 0.010$), but the two groups did not differ in neural responses to more trustworthy faces ($B = 0.08$, $SE = 0.15$, $P = 0.625$).

Discussion

Adolescents vary in their tendency to perceive others as trustworthy (Nowakowski et al., 2010), and attachment theory may contribute to our understanding of such individual differences. Advancing the literature on the contribution of early attachment to later SIP (Dykas and Cassidy, 2011), we compared trustworthiness evaluations of emotionally neutral faces between adolescents classified as secure and insecure in toddlerhood and assessed attachment-group differences in such evaluations at the behavioral and neural levels. Overall, our findings suggest that secure vs insecure attachment is associated with more open processing of facial trustworthiness, as evident in both adolescents' trustworthiness ratings and their neural activation across several ROIs.

At the behavioral level, adolescents from both attachment groups provided lower trustworthiness ratings as the consensus ratings decreased, suggesting that adolescents were able to discriminate varying levels of trustworthiness from unfamiliar, emotionally neutral faces. Consistent with our hypothesis,

the association between consensus and adolescents' ratings was weaker for adolescents with an insecure (*vs* secure) attachment history. Although both groups provided similar ratings of the more trustworthy faces, adolescents who had an insecure attachment history rated the less trustworthy faces as more trustworthy compared with their secure counterparts. These findings suggest that secure (*vs* insecure) attachment is associated with a greater capacity to openly identify aversive properties of untrustworthy faces. These findings also align with the proposition that secure (*vs* insecure) attachment is associated with an openness to process negative social information (Dykas and Cassidy, 2011), as well as past behavioral evidence that secure *vs* insecure youth are able to recall more correctly negative attachment-related information (e.g. details about separation events; Alexander *et al.*, 2010) and to attend more closely to mothers' faces characterized by positive and negative emotional valence (Vandevivere *et al.*, 2014). Similar to recalling memories of rejection and viewing negative emotions, untrustworthy faces may elicit experiences of social threat (e.g. Rubin *et al.*, 1998). Hence, the ability to openly process and tolerate such aversive social information may differ across attachment groups.

To understand neural processes underlying attachment-related differences in trustworthiness perception, we were guided by past investigations that have highlighted greater involvement of bilateral amygdala, bilateral fusiform, right AI and right pSTS in response to untrustworthy *vs* trustworthy faces (Winston *et al.*, 2002; Engell *et al.*, 2007; Bzdok *et al.*, 2011; Mende-Siedlecki *et al.*, 2013; Kragel *et al.*, 2015). Paralleling findings at the behavioral level, adolescents with a secure attachment history showed enhanced activation to decreases in trustworthiness, whereas those with an insecure attachment history showed attenuated activation to less trustworthy faces. Moreover, our preliminary analysis indicated that this differential neural tracking of trustworthiness by attachment group did not differ as a function of ROI. According to neural evidence on trustworthiness perception (e.g. Bzdok *et al.*, 2011; Kragel *et al.*, 2015), adolescents' increased activation in the bilateral amygdala, bilateral fusiform, right AI and right pSTS to less trustworthy faces (which may serve as threatening or unpleasant social cues) may indicate flexible engagement of brain regions involved in motivational relevance, social salience and mentalizing. Thus, our findings indicate that the secure group may be more capable of processing increasingly threatening or unpleasant facial stimuli, whereas the insecure group may fail to flexibly recruit relevant regions to support interpretation and representation of such faces (also see Vrtička *et al.*, 2014).

Together, our behavioral and neural findings consistently indicate attachment-related differences in trustworthiness evaluation, with secure (*vs* insecure) attachment associated with greater capacity to openly process less trustworthy faces. One potential explanation is that insecure individuals may engage in defensive exclusion (e.g. downplay the salience and refrain from thorough processing) of aversive social information to protect themselves from potential social pain (Dykas and Cassidy, 2011). Alternatively, individuals with a history of insecure attachment may simply be less able to distinguish among levels of trustworthiness cues in emotional neutral faces, leading to insufficient response to untrustworthy cues and poor performance on trustworthiness evaluation. Future research aimed at testing these competing mechanisms is needed.

Several limitations should be considered. Firstly, our participants were predominantly White, and replication studies among

more racially diverse samples are needed. Secondly, our sample size was modest, and there was non-negligible attrition from toddlerhood to adolescence. Nonetheless, missing data analyses indicated few differences between those who returned at the adolescent time point and those who did not, suggesting that our results are not likely to be biased by this attrition. Moreover, the neural findings paralleled the behavioral results, which increase confidence in the findings. Given the small sample, we were unable to examine whether associations differed as a function of the subcategories of insecure attachment (i.e. avoidant, ambivalent, controlling or insecure-other). Because distinct SIP patterns have been theorized to characterize the insecure subgroups at neural (e.g. Vrtička and Vuilleumier, 2012; Long *et al.*, 2020) and behavioral (e.g. Corriveau *et al.*, 2009) levels, future studies designed to oversample insecure subgroups are needed to obtain sufficient power to test potential differential SIP patterns as a function of insecure subgroups.

In conclusion, employing a multimethod longitudinal design that spanned 10 years, we provide unique evidence on the prospective association between early attachment and later SIP. Across observed behavioral and neural responses, the findings offer a valuable lens toward understanding the implications of early attachment relationships in trustworthiness evaluation during adolescence.

Acknowledgements

We are grateful to the families who participated in this research. We also thank Elissa Thomann Mitchell, Heather Ross, Jordan Bodway, Ethan McCormick, Michael Perino and Tae-Ho Lee for assistance with data collection and Bonnie Conley and Susan Paris who coded the modified Strange Situation assessments.

Funding

This study was supported by grants from the National Science Foundation, USA, (BCS 1539651) to NLM and EHT; the USDA National Institute of Food and Agriculture, USA, (ILLU-793-362) to NLM; the University of Illinois Research Board, USA, to NLM; and the Jacobs Foundation, Switzerland, (2014–1095) to EHT.

Conflict of interest

The authors have no conflict of interests to report.

References

- Adolphs, R., Tranel, D., Damasio, A.R. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition*, **52**(1), 61–9.
- Alexander, K.W., O'Hara, K.D., Bortfeld, H.V., Anderson, S.J., Newton, E.K., Kraft, R.H. (2010). Memory for emotional experiences in the context of attachment and social interaction style. *Cognitive Development*, **25**(4), 325–38.
- Beckmann, C.F., Smith, S.M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, **23**(2), 137–52.
- Bowlby, J. (1973). Attachment and loss: Volume II: Separation, anxiety and anger. In *Attachment and loss: Volume II: Separation, anxiety and anger*. London: The Hogarth press and the institute of psycho-analysis, 1–429.
- Bowlby, J. (1980). Attachment and loss: Volume III: Loss, sadness and depression. In *Attachment and Loss: Volume III: Loss, Sadness and Depression*. London: The Hogarth press and the institute of psycho-analysis, 1–462.

- Bretherton, I., Munholland, K.A. (2008). Internal working models in attachment relationships: elaborating a central construct in attachment theory. In: Cassidy, J., Shaver, P.R. editors. *Handbook of Attachment: Theory, Research, and Clinical Applications*, 2nd edn, New York: The Guilford Press, 102–27.
- Bzdok, D., Langner, R., Caspers, S., et al. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure & Function*, **215**(3–4), 209–23.
- Carlo, G., Randall, B., Rotenberg, K., Armenta, B. (2010). A friend in need is a friend indeed: Exploring the relations among trust beliefs, prosocial tendencies, and friendships. In: Rotenberg, K. editor, *Interpersonal Trust during Childhood and Adolescence*. Cambridge: Cambridge University Press, 270–94.
- Cassidy, J., Marvin, R.S., With The Macarthur Working Group. (1992). Attachment organization in preschool children: procedures and coding manual. *Unpublished Manuscript*, University of Virginia, 125–31.
- Corriveau, K.H., Harris, P.L., Meins, E., et al. (2009). Young children's trust in their mother's claims: longitudinal links with attachment security in infancy. *Child Development*, **80**(3), 750–61.
- Dykas, M.J., Cassidy, J. (2011). Attachment and the processing of social information across the life span: theory and evidence. *Psychological Bulletin*, **137**(1), 19–46.
- Engell, A.D., Haxby, J.V., Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, **19**(9), 1508–19.
- Escobar, M.J., Rivera-Rei, A., Decety, J., et al. (2013). Attachment patterns trigger differential neural signature of emotional processing in adolescents. *PLoS ONE*, **8**(8), e70247.
- Frith, U., Frith, C.D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **358**(1431), 459–73.
- Gorgolewski, K.J., Varoquaux, G., Rivera, G., et al. (2015). NeuroVault.Org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, **9**(APR), 8.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, **51**(1), 59–67.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, **17**(2), 825–41.
- Julian, J.B., Fedorenko, E., Webster, J., Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, **60**(4), 2357–64.
- Kammermeier, M., Duran Perez, L., König, L., Paulus, M. (2019). Attachment security and attention to facial emotional expressions in preschoolers: an eye-tracking study. *British Journal of Developmental Psychology*, **38**(2), 167–85.
- Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, **17**(11), 4302–11.
- Kragel, P.A., Zucker, N.L., Covington, V.E., LaBar, K.S. (2015). Developmental trajectories of cortical-subcortical interactions underlying the evaluation of trust in adolescence. *Social Cognitive and Affective Neuroscience*, **10**(2), 240–7.
- Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, **83**(404), 1198–202.
- Long, M., Verbeke, W., Ein-Dor, T., Vrtička, P. (2020). A functional neuro-anatomical model of human attachment (NAMA): insights from first- and second-person social neuroscience. *Cortex*, **126**, 281–321.
- Lundqvist, D., Flykt, A., Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, **91**(630), 2.
- McElwain, N.L., Holland, A.S., Engle, J.M., Wong, M.S. (2012). Child anger proneness moderates associations between child-mother attachment security and child behavior with mothers at 33 months. *Journal of Family Psychology*, **26**(1), 76–86.
- Mende-Siedlecki, P., Said, C.P., Todorov, A. (2013). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, **8**(3), 285–99.
- Menon, V., Uddin, L.Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure & Function*, **214**(5–6), 655–67.
- Muthén, L.K., Muthén, B.O. (2017). *Mplus User's Guide*, 8th edn, Los Angeles, CA: Muthén & Muthén.
- Network. (2001). Child-care and family predictors of preschool attachment and stability from infancy. *Developmental Psychology*, **37**(6), 847–62.
- Nowakowski, M.E., Vaillancourt, T., Schmidt, L.A. (2010). Neurobiology of interpersonal trust. In: Rotenberg, K.J. editor. *Interpersonal Trust during Childhood and Adolescence*. Cambridge: Cambridge University Press, 28–55.
- Ravindran, N., Hu, Y., McElwain, N.L., Telzer, E.H. (2020). Dynamics of mother-adolescent and father-adolescent autonomy and control during a conflict discussion task. *Journal of Family Psychology*, **34**(3), 312–21.
- Rotenberg, K.J., MacDonald, K.J., King, E.V. (2004). The relationship between loneliness and interpersonal trust during middle childhood. *Journal of Genetic Psychology*, **165**(3), 233–49.
- Rotenberg, K.J., Boulton, M.J., Fox, C.L. (2005). Cross-sectional and longitudinal relations among children's trust beliefs, psychological maladjustment and social relationships: are very high as well as very low trusting children at risk? *Journal of Abnormal Child Psychology*, **33**(5), 595–610.
- Rotter, J.B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, **26**(5), 443–52.
- Rubin, K.H., Bukowski, W., Parker, J.G. (1998). Peer interactions, relationships, and groups. In: Damon, W., (Series Ed.) & Eisenberg, N. (Vol. Ed.), *Handbook of Child Psychology: Social, Emotional, and Personality Development*, 5th edn, Vol. 3, New York: Wiley, 619–700.
- Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, **42**(11), 1435–46.
- Schore, A.N. (2001). Effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Mental Health Journal*, **22**(1–2), 7–66.
- Simpson, J.A. (2007). Psychological foundations of trust. *Current Directions in Psychological Science*, **16**(5), 264–8.
- Smith, S.M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, **17**(3), 143–55.
- Solomon, J., George, C. (2008). The measurement of attachment security and related constructs in infancy and early childhood. In: *Handbook of Attachment: Theory, Research, and Clinical Applications*, 2nd edn, New York: The Guilford Press, 383–416.
- Sroufe, L.A., Fleeson, J. (1986). Attachment and the construction of relationships. In: Hartup, W.W., Rubin, Z. editors. *Relationships and Development*. Hillsdale, NJ: Erlbaum, 51–71.
- Todorov, A., Pakrashi, M., Oosterhof, N.N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, **27**(6), 813–33.

- Tohka, J., Foerde, K., Aron, A.R., Tom, S.M., Toga, A.W., Poldrack, R.A. (2008). Automatic independent component labeling for artifact removal in fMRI. *NeuroImage*, **39**(3), 1227–45.
- Vandevivere, E., Braet, C., Bosmans, G., Mueller, S.C., De Raedt, R. (2014). Attachment and children's biased attentional processing: evidence for the exclusion of attachment-related information. *PLoS ONE*, **9**(7), e103476.
- Vrtička, P., Sander, D., Anderson, B., Badoud, D., Eliez, S., Debbané, M. (2014). Social feedback processing from early to late adolescence: influence of sex, age, and attachment style. *Brain and Behavior*, **4**(5), 703–20.
- Vrtička, P., Vuilleumier, P. (2012). Neuroscience of human social interactions and adult attachment style. *Frontiers in Human Neuroscience*, **6**(July), 1–17.
- Wentzel, K.R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, **62**(5), 1066–78.
- Willis, J., Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, **17**(7), 592–8.
- Winston, J.S., Strange, B.A., O'Doherty, J., Dolan, R.J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, **5**(3), 277–83.
- Zimmermann, P., Iwanski, A. (2015). Attachment in middle childhood: associations with information processing. *New Directions for Child and Adolescent Development*, **2015**(148), 47–61.