



**HAL**  
open science

## Age-related neural correlates of facial trustworthiness detection during economic interaction.

Emilie Salvia, Katell Mevel, Grégoire Borst, Nicolas Poirer, Grégory Simon, François Orliac, Olivier M Etard, Astrid Hopfensitz, Olivier Houdé, Jean-François Bonnefon, et al.

### ► To cite this version:

Emilie Salvia, Katell Mevel, Grégoire Borst, Nicolas Poirer, Grégory Simon, et al.. Age-related neural correlates of facial trustworthiness detection during economic interaction.. *Journal of Neuroscience, Psychology, and Economics*, 2020, 13 (1), pp.19-33. 10.1037/npe0000112 . hal-03025378

**HAL Id: hal-03025378**

**<https://hal.science/hal-03025378>**

Submitted on 26 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **AGE-RELATED NEURAL CORRELATES OF FACIAL TRUSTWORTHINESS DETECTION DURING ECONOMIC INTERACTION**

+Emilie Salvia<sup>1,2</sup>, +Katell Mevel<sup>1,2</sup>, Grégoire Borst<sup>1,2,3</sup>, Nicolas Poirel<sup>1,2,3</sup>, Grégory Simon<sup>1,4</sup>, François Orliac<sup>1</sup>, Olivier Etard<sup>5</sup>, Astrid Hopfensitz<sup>6</sup>, Olivier Houdé<sup>1,2,3</sup>, Jean-François Bonnefon<sup>7</sup>, \*Wim De Neys<sup>1,2</sup>

+both authors contributed equally

1 Laboratory for the Psychology of Child Development and Education, Sorbonne, CNRS UMR 8240, Paris, France

2 Université de Paris, Sorbonne Paris Cité, Paris, France,

3 Institut Universitaire de France, Paris, France,

4 Université de Caen Normandie, Caen, France

5 Centre Hospitalier Universitaire de Caen, Service des Explorations Fonctionnelles du Système Nerveux, Caen, France

6 Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France

7 Toulouse School of Economics, Center for Research in Management (CNRS UMR5303), University of Toulouse Capitole, Toulouse, France

\*Corresponding author:

Wim De Neys

Sorbonne – Labo Binet

Paris Descartes University

46 Rue Saint-Jacques

75005 Paris

France

Email: [wim.de-neys@parisdescartes.fr](mailto:wim.de-neys@parisdescartes.fr)

*(In Press). Journal of Neuroscience, Psychology, and Economics.*

## **AGE-RELATED NEURAL CORRELATES OF FACIAL TRUSTWORTHINESS DETECTION DURING ECONOMIC INTERACTION**

Some economic transactions require people to trust strangers, whose trustworthiness is unknown. In these circumstances, behavioral studies have shown that adults (but not young adolescents) seem to have some minimal ability to detect the trustworthiness of adult strangers based on their facial features. In this study we explored the neural correlates of this facial trustworthiness detection. A group of adolescents and adults played a series of economic trust games with adult trustees of which we had previously recorded the strategy. Results showed that when adult investors were looking at the picture of a trust abusing trustee, the left amygdala was relatively more activated than when they were looking at a trust honoring player. Younger adolescents did not show this pattern and responded with a more pronounced deactivation when facing a trust abusing trustee. An exploratory whole brain analysis detected a similar age shift for mentalizing regions of the brain. Our results fit with an emerging model suggesting that the amygdala is implicated in an associative learning process which progressively refines a mapping of faces onto trustworthy behavior, and may result in avoiding to be exploited by untrustworthy strangers.

Keywords: trust game; amygdala; fMRI; trustworthiness detection

## **AGE-RELATED NEURAL CORRELATES OF FACIAL TRUSTWORTHINESS DETECTION DURING ECONOMIC INTERACTION**

The human species shows a remarkable ability to trust and cooperate (Thielman & Hilbig, 2015; Tomasello, 2009). This ability to trust is considered to be one of the main reasons behind the success of human societies. But while trusting others has many potential benefits, our trust can also be abused and exploited. Consequently, it can be beneficial for humans to accurately judge the trustworthiness of others.

Available evidence suggests that people have a strong tendency to judge the trustworthiness of their interaction partners on the basis of their facial appearance (Adolphs & Tusche, 2017; van 't Wout & Sanfey, 2008; Todorov, 2017). Put differently, we tend to trust people who “look” trustworthy. Research indicates that these facial trustworthiness judgments show very high inter-rater agreement and are made effortlessly in a mere 100 ms or less (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006). Neuroscientific studies have pointed to a critical role of the amygdala in this process (e.g., Adolphs, Tranel, & Damasio, 1998; Santos, Almeida, Oliveiros, & Castelo-Branco, 2016).

In general, the amygdala has long been connected with basic emotional processing and threat detection (Pessoa & Adolphs, 2011). Numerous fMRI studies on trustworthiness evaluation found that the amygdala shows an increased activation when participants are presented with untrustworthy looking face stimuli (e.g., Bzdok et al., 2011; Engell, Haxby, & Todorov, 2007; Todorov, Mende-Siedlecki, & Dotsch, 2013; Mende-Siedlecki, Said, & Todorov, 2013; Freeman, Stolier, Ingbreetsen, & Hehman, 2014; Rule, Krendl, Ivcevic, & Ambady, 2013; Winston, Strange, O’Doherty, & Dolan, 2002). This increased response to untrustworthy faces is sometimes described as following a linear trend while others have found a quadratic response pattern with higher responses at the extremes of the trustworthiness dimension (Santos et al., 2016). In addition, lesion studies indicate that patients with amygdala damage judge faces to look abnormally trustworthy and approachable (Adolphs & Tusche, 2017; Adolphs et al., 1998).

While it is well established that people readily make facial trustworthiness judgments, it is also clear that these judgments are rarely accurate. That is, people who look untrustworthy do not necessarily behave untrustworthily (Efferson & Vogt, 2013; Olivola, Funk, & Todorov, 2014; Todorov, Olivola, Dotsch, Mende-Siedlecki, 2014; Wilson & Rule, 2015; but see also Little, Jones, DeBruine, & Dunbar, 2013; Tognetti, Berticat, Raymond, & Faurie, 2013). For example, Rule et al. (2013) created an

experimental setting in which participants could cheat on a test or not. When a different group of participants were shown pictures of the cheating and non-cheating participants, their ratings showed high agreement on who looked trustworthy and who did not. However, the pictures of people who actually cheated on the test were not rated differently than the pictures of people who did not. Critically, in an fMRI study with the same pictures, Rule et al. found that the amygdala activation for the pictures of cheating and non-cheating individuals did also not differ. This suggests that the amygdala activation tracks perceived trustworthiness but does not provide an accurate (i.e., so-called “honest”) trust signal.

Taken together, the available evidence indicates that people’s trustworthiness ratings of faces are generally inaccurate. However, recent behavioral work also indicates that under specific conditions people’s trust decisions can retain some minimal but observable accuracy (e.g., Bonnefon, Hopfensitz, & De Neys, 2013; Stirrat & Perret, 2010; see Bonnefon, Hopfensitz, & De Neys, 2017 for review). These specific conditions concern cases in which people evaluate faces in the context of an incentivized Trust Game (Berg, Dickhaut, & McCabe, 1995). The Trust Game is a popular behavioral economic paradigm that allows researchers to capture trusting decisions in a controlled but reasonable realistic economic setting. This simple game involves two players, A and B. Player A, the Investor or Trustor, is endowed with an initial amount of money (e.g., \$10) and can choose to invest this money and transfer it to player B, the Trustee. If player A decides to transfer their money, the experimenter multiplies the amount by a factor larger than 1 (typically 3) before giving it to player B. Player B then decides how much of this enlarged endowment (e.g., \$30), if any, they would like to return to player A. Both players are completely informed about the procedure but cannot communicate during the game. In essence, in this game the Investor needs to decide whether or not to trust the Trustee, and the Trustee has the option between honoring and abusing trust. If the Investor transfers their endowment and the Trustee honors this trust by returning more money than originally transferred, the Investor will make a nice profit and will end up with more money than the initial \$10. However, if the Trustee abuses trust and returns no or only little money, the Investor ends up with less money than their initial endowment.

In a set of studies Bonnefon et al. (2013) first asked a group of participants to play the role of Trustee in the Trust game. Participants were fully instructed about the structure of the game and were asked what they would do if the Investor transferred their endowment: Would they keep all of the money or split it equally and return half of it? After the experiment, Trustees were then paired with a random Investor and paid depending on their decision. Hence, this initial phase is giving us an operationalization of the actual trustworthiness of the Trustee: We know whether they abused the

Investor's trust and kept all of the money, or honored trust and split the money. In the second stage of the study, another group of participants played several Trust game rounds in the role of Investor. Before each round they were shown a picture of the Trustee (taken from the first phase of the study) and decided whether or not they wanted to transfer their endowment to the Trustee. One of the rounds was randomly selected and the Investors received whatever money they made in that round. Hence, both the Investor and Trustees were playing for real monetary pay-offs. The critical observation was that participants transferred more frequently (i.e., a 5% to 7% transfer increase) to Trustees who honored trust than to Trustees who abused trust. This indicates that people have some minimal accuracy in reading trustworthiness from the faces of unknown adults when they play the Trust Game (see also De Neys, Hopfensitz, & Bonnefon, 2013, 2017; Stirrat & Perret, 2010).

Here we report the first fMRI study that looks at the neural basis of this accurate facial trustworthiness detection in the Trust Game. In the scanner our participants played a series of trust games as Investor with Trustees whose strategy had been previously recorded. Participants were shown the picture of the Trustee before making a transfer decision. We contrasted the neural activation when participants saw pictures of Trustees who abused and honored trust. We were especially interested in the possible role of the amygdala in this trustworthiness signaling: Does amygdala activation indicate we're looking at an abuser when deciding to trust them with our money? This would suggest that the amygdala response can serve as an honest trust signal in the context of the Trust Game.

Critically, we tested both a group of young adults (age 19-26) and young adolescents (age 11-16). In general, amygdala responsiveness is known to show important age-related development (Gee et al., 2013; Tottenham & Gabard-Durnam, 2017; Zebrowitz, Ward, Boshyan, Gutchess, & Hadjikhani, 2018). More directly, previous behavioral work with the Trust Game has indicated that adolescents are less accurate in detecting trustworthiness from the face than adults (De Neys, Hopfensitz, & Bonnefon, 2015). De Neys et al. had 540 adolescents aged between 13 and 18 years decide whether or not to trust an unknown adult based on their picture in a number of Trust Game rounds. Results showed that trustworthiness detection accuracy increased with age: younger adolescents' transfer decisions differentiated less well between trust honoring and abusing Trustees than those of later adolescents and young adults. We therefore expected that if the amygdala (or any other brain region) showed sensitivity to the Trustees' trustworthiness status, the activation pattern would differ in the adolescent and adult groups.

## METHOD

## Participants

Forty-nine (23 adults and 26 adolescents) participants took part in this study. Among the 23 scanned adults, data from one participant who exhibited a within-run maximal amplitude of translational or rotational between-volumes displacement above 2 mm or 1.5 degree, respectively, was discarded from the analysis. A second adult had to be excluded due to technical scanner problems. Among the 26 scanned adolescents, three had to be discarded because of within-run displacements above 2mm translation or 1.5 degrees rotation, four had to be excluded due to technical scanner problems, and one had to be excluded because of a neurological abnormality. Consequently, datasets of 21 adults (mean age  $\pm$  21.4, SD  $\pm$  1.9 years, aged from 19.1 to 25.9 years, 11 females) and 18 adolescents (mean age  $\pm$  13.3, SD  $\pm$  1.3 years, aged from 11.4 to 15.8 years, 10 females) were included in the final analysis. All participants were right handed, native French speakers. An informed consent form was signed by themselves or their parents/guardians (for minors). Each participant received a 25 euro gift voucher for their participation. The study was approved by the research ethics committee (CPP Nord Ouest III; ID RCB: 2014-A00935-42) and conformed to the 1964 Declaration of Helsinki.

## Material and Procedure

### Behavioral task design

*Trust Game.* Participants were familiarized with the rules of the Trust game before the scanning session. In the scanner, each participant played a total of 60 rounds in the role of Investor, with 60 different Trustees. Note that while the words 'Investor' and 'Trustee' are used for clarity here, they were never actually used in the experiment. The Trustee was simply called 'Player B'. Each round had the same structure: Participants were endowed with a sum of 4 euros and had to decide whether to keep the endowment, or to transfer that endowment to a Trustee, whose picture appeared on the screen. In case the endowment was transferred, it was multiplied by three, and the Trustee had to decide whether to keep the whole 12 euros, to return 6 euros to the Investor, or to return 4 euros to the Investor. We refer to these strategies as the Abuser, Cooperator, and Neutral strategies, respectively. These terms were not mentioned to the participants. The participants were informed that each Trustee had already recorded his or her strategy. Participants were also informed that one round would be randomly

selected after the study, and that the amount of money they made in that round would be added to the voucher they would receive at the end of the study.

Trustors' payoffs were based on a pairing with a randomly selected Trustee. Trustees did not receive any payoff in the current set of studies (they did receive their payoff in the previous study in which their picture had been taken and their strategy recorded, see Centorrino, Djemai, Hopfensitz, Milinski, & Seabright, 2015). In addition, because of possible ethical objections to a performance-dependent reimbursement with younger participants (e.g., De Neys et al., 2015), all Investors were paid the maximum payoff which was included in their voucher fee. Investors were not informed about their actual performance. Hence, our task design implied the use of deception in that Trustors believed that the Trustee's pay-off was contingent on the Trustor's own behavior and Trustors were not informed about their actual performance. Note also that in our variant of the Trust game the Trustee received a fixed show-up fee but no additional endowment in the game (e.g., see Glaeser, Laibson, Scheinkman, & Soutter, 2000). Trustors were informed about this feature. This can affect the absolute level of trust (e.g., due to inequality aversion Trustors may be more likely to transfer money overall, e.g., Ciriolo, 2007). However, given that our core interest lies in the accuracy of trustworthiness detection (i.e., relative contrast Abusers vs Cooperators) this design feature should not bias results (De Neys et al., 2017).

*Trustee pictures.* The Trustees shown to the participants had recorded their strategy and were incentivized in the context of a previous study (Centorrino et al., 2015). In this initial study, 84 young adults played the role of Trustee and recorded a movie introducing themselves. From each of these movies, a research assistant blind to the strategies of the Trustees extracted one frame in which the Trustee had the most neutral expression. Each picture was then cropped (left and right facial boundaries, chin and top of the eyebrows) to minimize display of clothing or hairstyle, and turned to black and white (Figure 1). The trustworthiness detection study of Bonnefon et al. (2013) used a set of 60 of these pictures selected so that the proportion of abusers and cooperators would be maximally similar for male Trustees (4 abusers, 18 cooperators) and female Trustees (3 abusers, 17 cooperators). Bonnefon et al. also avoided pictures of distinctively non-Caucasian Trustees to maximize homogeneity of the picture set. The same set of 60 pictures from Bonnefon et al. was also used for the present study. Consequently, there were a total of 7 pictures of Abusers, 35 pictures of Cooperators, and 18 pictures of Neutral players. Note that the proportion of cooperators in the 60-picture set is similar to the



proportion of cooperators usually found in trust game studies and the natural environment (e.g., Johnson & Mislin, 2011; Van Lange, 2015).

#### fMRI event-related design

Each trial started with the presentation of the number of the round (e.g., “Round 1”, 1000 ms) that was going to be played. Next, a picture of the Trustee was presented on the screen for 2000 ms. Subsequently, participants had to indicate whether they wanted to transfer money to the Trustee by pressing one of two buttons, after which they moved on to the next round, without receiving feedback about the strategy of the Trustee. Participants were informed that they had a maximum of 3000 ms to give their response (i.e. transfer money or keep it). As soon as they answered, a fixation cross was presented in the middle of the screen for periods ranging from 2200 ms to 6400 ms. The fixation cross was followed by the next trial. Trustee pictures were presented in a jittered pseudo-randomized order for each participant. The category sequence structure was identical for each participant (e.g. Cooperator picture, Cooperator picture, Abuser picture, Neutral picture, ... ) but the Trustee picture presented for each category was randomly selected among the available stimuli. Ten null events consisting of a fixation cross for a duration ranging from 2200 to 6400ms were also added to the design to optimize signal detection. The design was overall optimized using the Genetic Algorithm toolbox (Wager and Nichols, 2003). Figure 1 illustrates the trial sequence.

#### fMRI data acquisition

fMRI data were acquired at Cycleron (Caen, France) using a 3T scanner (Philips Achieva, Eindhoven, The Netherlands). Scanning was done as part of a larger one hour test session in which participants were also presented with an unrelated reasoning task. For each participant, the MR data were collected using planes parallel to the anterior commissure – posterior commissure line. fMRI time series were collected using T2\*-weighted FFE echo-planar imaging (EPI) sequence (31 axial slices; 3.75 mm thickness; no gap; reconstruction matrix = 640 x 640 mm; field of view = 240; repetition time = 2000 ms; echo time = 35 ms; flip angle = 80°; interleaved: bottom – up). The FOV covered the top of the cortex down to at least the base of the cerebellum. The first 6 volumes were discarded from each functional run to account for spin saturation effects. A high resolution T1-weighted structural and a T2\* non-EPI volumes were also acquired using 3D TFE (180 sagittal slices; resolution: 1 mm<sup>3</sup>; no gap; reconstruction matrix = 256 x 256 mm; field of view = 256; repetition time = 20 ms; echo time = 46 ms)

and 2D FFE (70 axial slices; resolution: 2 mm<sup>3</sup>; no gap; reconstruction matrix = 256 x 256 mm; field of view = 256; repetition time = 20 ms; echo time = 46 ms) sequences, respectively.

#### Fmri data analysis

Image pre-processing and statistical analyses were performed using Matlab 2014b (Mathworks Inc., Natick, USA) and SPM12 (Wellcome Department of Imaging Neuroscience, London, UK).

*Pre-processing.* fMRI datasets were preprocessed to optimize the anatomical mapping of the functional results using approaches inspired by Villain et al. (2010). More precisely, individual fMRI time-series were (1) slice time corrected, (2) realigned to correct for head movement, (3) co-registered to the respective individual T2\* volume, which was already previously co-registered to the high-resolution structural image T1-weighted volume, (4) transformed into the MNI152 T1 template space using the parameters derived from the spatial normalization of the T1-weighted image in DARTEL (Ashburner, 2007), (5) resampled to 2-mm isotropic voxel size and (6) smoothed with a 8-mm\_FWHM Gaussian Kernel. An explicit mask was designed, based on the averaged normalized segmented T1 data, to exclude non-gray matter voxels from further fMRI analyses.

*fMRI modeling and statistical analyses.* We analyzed the data voxel-wise using the general linear model (GLM) approach (Friston et al., 1995). For completeness, we modelled activation both during the critical face presentation phase in which participants saw the Trustees picture and during the subsequent transfer response phase during which they indicated their transfer decision. We defined six regressors: (1) a regressor, for each condition (i.e., the three faces types), related to the face presentation phase (i.e., the presentation of Neutral, Cooperator and Abuser faces) and (2) a regressor, for each condition (i.e., the three face types), related to the transfer phase (i.e., phase during which the participants indicated whether or not they wanted to transfer their money after the presentation of a Neutral, Cooperator and Abuser face). For each regressor we modelled the stimulus-evoked neural response as a boxcar function, with each event starting at face presentation or at transfer phase onset, and lasting 2000 ms or the time the participant took to make their transfer decision, respectively. These boxcar functions were convolved with the default canonical hemodynamic response function of SPM12. In addition, we included one constant term and the six realignment parameters (3 translations, 3 rotations) as covariates into our model, to account for changes in signal level and influence of head

motion on BOLD signal, respectively. We also applied a high-pass filter (cut off period = 128 s) to remove low-frequency drifts.

We first estimated the model at the individual level yielding two parameter estimates per condition (i.e., face presentation for Neutral, Cooperator and Abuser conditions; transfer for Neutral, Cooperator and Abuser conditions). From these values, we computed, for each participant and for both face presentation and transfer, the following statistical contrast maps: Abuser > Cooperator, Abuser > Neutral, Cooperator > Neutral. Our main interest was whether there was differential activation for Abuser and Cooperator faces. Note that some previous fMRI studies on trustworthiness evaluation pointed to a possible quadratic response pattern in the amygdala (i.e., no difference between trustworthy and untrustworthy faces per se, but increased activation for both untrustworthy and trustworthy vs neutral rated faces, e.g., Santos et al., 2016; Todorov et al., 2013). The additional Neutral contrasts would allow us to track such a pattern.

At the group level, we conducted two-sample t-tests, for each contrast described above, to test for activation differences in the adult and adolescent group (Adults > Adolescents) and track the hypothesized effect of age. In addition, we also performed supplementary analyses (see results) in which we a) combined adults and adolescents in one single group and performed one-sample t-tests to test correlations between amygdala BOLD signal and age, and b) test correlations between amygdala BOLD signal and behavioral data (i.e., the difference of average transfer rate between abusers and cooperators) in both age groups separately and combined.

We determined a priori bilateral amygdala regions of interest (ROI) based on the work of Rule et al. (2013). The ROIs were 4mm radius spheres centered on the voxels that showed peak activation within the left and right amygdala in Rule et al. (2013): left amygdala [-24, 0, -12], right amygdala [27, -3, -18]. Figure 2 depicts the exact ROI location. We first performed analyses using the amygdala ROIs for which we applied both voxel-level and cluster-level thresholds at p FWE corrected < .05. For subsequent whole brain analyses, we applied a voxel-level threshold at  $p < .001$  (uncorrected for multiple comparisons) and cluster-level threshold at  $p < .05$  (FWE corrected).

## RESULTS

### *Behavioral results*

For each participant we calculated the average transfer rate to abusers, cooperators, and neutral Trustees. These averages were subjected to a 3 (Trustee Strategy, within-subjects; abuser,

cooperator, or neutral) x 2 (Age Group, between-subjects; adolescent or adult) mixed model analysis of variance (ANOVA). Table 1 gives an overview of the results. As Table 1 indicates, behavioral performance was fairly similar across all conditions. ANOVA results confirmed that the transfer rates did not significantly differ as a function of the Trustee Strategy factor,  $F(2, 111) = 0.718$ ,  $p = .490$ ,  $\eta_p^2 = 0.013$ , Age Group factor,  $F(1, 111) = 0.418$ ,  $p = .519$ ,  $\eta_p^2 = 0.004$ , or their interaction,  $F(2, 111) = 0.160$ ,  $p = .852$ ,  $\eta_p^2 = 0.003$ . This indicates that our participants' behavioral performance did not show the previously observed trustworthiness detection effects. This was not unexpected given that the previous behavioral studies reported small effect sizes and tested much larger samples (e.g.,  $n = 208$  in Bonnefon et al., 2013, Study 1; and  $n = 540$  in De Neys et al., 2015). Our goal in the present study was to investigate a potentially more pronounced neural signature.

### *fMRI results - face presentation phase*

We start by focusing on the amygdala ROI and afterwards move on to a whole-brain analysis. For all analyses we contrasted how the activation in the adult group differed from the adolescent group (Adults > Adolescents) to track the hypothesized effect of age. For ROI analyses, voxel-level and cluster-level thresholds were set at  $p < .05$  (FWE corrected). For whole brain analyses, voxel-level and cluster-level thresholds were set at  $p < .001$ , (uncorrected for multiple comparisons) and  $p < .05$  (FWE corrected), respectively

*Amygdala ROI analyses.* The Abuser > Cooperator contrast revealed greater activation within the left amygdala (left amygdala: [-26, 2, -14];  $k = 3$ ,  $t = 3.47$ ,  $p < .05$ , FWE; see Figure 3A) for adults than for adolescents whereas no significant difference was observed in the right amygdala. Interestingly, the Abuser > Neutral contrast showed the same pattern with greater activation within the left amygdala (left amygdala: [-26, 2, -14];  $k = 7$ ,  $t = 3.77$ ,  $p < .05$ , FWE; see Figure 3A) for adults than for adolescents and no significant activation difference within the right amygdala. The Cooperator > Neutral contrast did not show any significant activation differences.

Figure 3B plots the activation pattern in the left amygdala in all conditions and both age groups to further interpret the findings. As the figure indicates, adults show less deactivation when looking at trust abusers than at cooperators and neutral players resulting in a relative activation increase in the Abuser condition. With adolescents we find the opposite pattern as they show a more pronounced

deactivation in the Abuser condition. Taken together, these findings indicate that there is an age related amygdala activation increase when we face Trustees who will abuse our trust.

*Supplementary ROI analyses.* In our main analyses, we looked at the impact of age by contrasting activation in the adult and adolescent group. Since there was some age variability within each age group, we also entered age as a continuous predictor in our analyses and tested whether it correlated with the observed left amygdala ROI activation. Results showed that this was the case: both for the Abuser > Cooperator ( $k = 1$ ,  $t = 3.66$ ,  $p < .05$ , FWE) and Abusers > Neutral ( $k = 5$ ,  $t = 3.71$ ,  $p < .05$ , FWE) contrast there was a positive correlation between age and activation of the left amygdala ROI. The scatterplot in Figure 4 illustrates the findings.

One might note that the Figure 4 scatterplot also suggests that two adolescents present very low amygdala contrast estimates which might be driving the positive age correlations. We thus redid the analyses after removing possible outliers from the data (i.e., removed data beyond  $+ / - 2$  standard deviations from the mean). Results indicated that the findings were robust. Even when the two possible outliers were excluded, we still observe a positive correlation between age and activation within the left amygdala ROI for both the Abuser > Cooperator ( $k = 1$ ,  $t = 3.36$ ,  $p < .05$ , FWE) and the Abuser > Neutral ( $k = 1$ ,  $t = 3.43$ ,  $p < .05$ , FWE) contrasts.

Finally, we also explored whether the individual adolescents' and adults' left amygdala ROI activation for the critical Abuser > Cooperator contrast correlated with their behavioral performance (i.e., average transfer rate to abusers – average transfer rate to Cooperators). However, none of the correlations reached significance (Adults:  $r = .07$ ,  $p = .75$ ; Adolescents:  $r = .23$ ,  $p = .37$ ; Adults+Adolescents:  $r = .17$ ,  $p = .29$ ; all  $p$  FWE, see Figure 5). This again indicates that behavioral effects are subtle. The more pronounced activation of the left amygdala when faced with abusers did not result in a decreased transfer to these Trustees.

*Whole-brain analysis.* In addition to our amygdala ROI contrasts we also ran explorative whole-brain analyses. As with the main ROI analyses, we contrasted how the activation in the adult group differed from the adolescent group (Adults > Adolescents) in our contrasts of interest. Table 2 gives a full overview of the results. As the table indicates, in addition to the amygdala, a number of additional parietal (e.g., Angular Gyrus, Inferior Parietal Lobule, Precuneus), temporal (e.g., Inferior Temporal Gyrus), and frontal regions (e.g., Medial Frontal Gyrus) also show an age related increase when facing trust abusers. Although post-hoc and speculative, we note that many of these regions have been

implicated in Theory of Mind and perspective taking tasks (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009). We come back to this point in the discussion section.

#### *fMRI results - Transfer response phase*

For completeness, we modelled activation both during the critical face presentation phase in which participants saw the Trustees picture and the subsequent transfer response phase during which they indicated their transfer decision. During the transfer phase neither the amygdala ROI analyses nor the whole brain analyses showed significant activation effects in any of our contrasts of interest.

### **DISCUSSION**

In this study we explored the neural basis of facial trustworthiness detection in the Trust Game. We had a group of adolescents and adults play a series of trust games as Investor with Trustees of which we had previously recorded the strategy. Results showed that when adult Investors were looking at the picture of a trust abusing Trustee, the left amygdala was relatively more activated than when they were looking at a trust honoring or neutral player. Younger adolescents did not show this pattern and responded with a more pronounced deactivation when facing a trust abusing Trustee. This pattern suggests that the previously observed behavioral age-related increase in Trust Game transfer accuracy and accurate trustworthiness detection among adults (Bonnefon et al., 2013; De Neys et al., 2015) might be mediated by the amygdala. Thereby, the results suggest that in the specific context of the Trust Game the amygdala activation might help to accurately track the trustworthiness of our interaction partners.

Our exploratory whole-brain analyses indicated that in addition to our amygdala ROI various parietal, temporal, and frontal regions also showed age related modulation when facing trust Abusers. One of the additionally activated regions, the angular gyrus, might be especially interesting. This region and the adjacent temporal parietal junction (TPJ) are often implicated in Theory of Mind and perspective taking tasks (Schurz et al., 2014; Van Overwalle, 2009). Critically, imaging studies with various economic games also point to a role of the TPJ in contemplating altruistic or selfish decision options (Morishima, Schunk, Bruhin, Ruff, & Fehr, 2012; van den Bos, van Dijk, Westenberg, Rombouts, & Crone, 2011). This TPJ activation is known to show age related differentiation (Fett, Gromann, Giampietro, Shergill, & Krabbendam, 2012; van den Bos et al., 2011). More specifically, age related TPJ activation increases

have been found to be especially prominent when dealing with negative social signals (e.g., when getting feedback about Trustees' defection, Fett et al., 2012). In this light, the fact that the angular gyrus/TPJ shows an age related activation increase when being faced with Abusers<sup>1</sup> in the current study might suggest that perspective taking is specifically recruited when adults are dealing with potential trust abuse.

We want to stress that some caution is needed in interpreting the present findings. Our study is but the first in which the neural basis of accurate face based trustworthiness detection in a Trust Game is explored. Obviously, the result will need to be replicated before drawing strong conclusions (Nosek, Spies, & Motyl, 2012). But even then one should be wary of possible over-interpretation. As we noted, previous large scale behavior studies already indicated that although Investors perform slightly better than chance (e.g., 6% transfer increase to cooperators vs abusers) their decisions are far from perfect (Bonnefon et al., 2013; De Neys et al., 2017). In the present small scale study we found that the increased amygdala response to untrustworthy Trustees did not result in an observable behavioral effect. The point is that although we believe it is important to try to identify the neural basis of a presumably noisy and coarse facial trust detection process, such a neural substrate should not be taken to imply that relying on facial trust cues (or associated brain activation) to make trust decisions is wise (McCabe & Castel, 2008). A key issue is that trust abuse in modern day societies is rare (e.g., Van Lange, 2015). For example, in our current sample, only 12% of Trustees decided to abuse trust and keep all of the transferred money. Hence, simply because the base rate of cooperation is so high, deciding to (dis)trust on the basis of an imperfect face-based signal can be counterproductive. By not transferring to those misclassified as abusers, we are missing out on the trust bonus. Consequently, a simple strategy to trust indiscriminately can be shown to result in higher payoffs than relying on imperfect trustworthiness detection cues (e.g., De Neys et al., 2017; Todorov, Funk, Olivola, 2015).

The present findings give rise to a number of further questions. For example, one might wonder about the discrepancy between the present results and those of Rule et al. (2013) who found that amygdala activation did not accurately track people's cheating behavior in a mock test. As with behavioral trust studies, one possible reason for the discrepancy concerns the incentivized Trust Game task context which has been argued to create favorable conditions for trustworthiness detection (Bonnefon et al., 2017). Interestingly, in another task context and with a different type of evaluation – perceptions of leadership – Rule and collaborators did find that face-based first impressions can be accurately reflected in amygdala activation (Rule, Moran, Freeman, Whitfield-Gabrieli, Gabrieli, &

---

<sup>1</sup> See Figure 6 for a detailed overview of the angular gyrus/TPJ pattern of activation in all our study conditions.

Ambady, 2011). Rule et al. (2011) presented participants with pictures of CEOs in the fMRI scanner. Prior behavioral work (Rule & Ambady, 2008) had already established that people's judgments of leadership on the basis of these pictures, corresponded to an objective measure of success: perceived leadership accurately predicted the CEOs companies' profits. Critically, Rule et al. (2011) observed that this was also accurately reflected in the amygdala activation: the amygdala was more responsive when viewing faces of CEOs whose companies had higher profits. Hence, the accuracy of first impressions reflected by amygdala activation may depend on the nature of the impression and task context.

Another issue concerns the possible origin of accurate face-based trust impressions. Where does the (minimal) accuracy come from? One speculative explanation involves the role of stimulus generalization or learning mechanisms (FeldmanHall et al., 2018; Over & Cook, 2018). For example, according to Over and Cook's Trait-Inference-Mapping framework, trait inferences are products of mappings between locations in 'face space' and 'trait space'. The idea is that when one repeatedly encounters individuals with particular facial features who subsequently exhibit certain behavior, a mapping forms between the corresponding face and trait representation. Once acquired, these mappings mediate spontaneous trait judgments from faces: when the face of a stranger falls close to a mapped location in face space, the associated trait will be automatically activated (Over & Cook, 2018). Interestingly, recent fMRI evidence also points to a role of the amygdala in the learned avoidance of untrustworthy behaving partners in the Trust Game (FeldmanHall et al, 2018). In their study FeldmanHall et al. had participants play repeatedly with the same Trustees. The Trustees were instructed to behave more or less trustworthily (i.e., return money more frequently). In a second stage, the participants played with a new set of Trustees whose pictures were morphed to resemble the faces of the trustworthy and untrustworthy Trustees from the initial stage. Results showed that participants were less likely to transfer money to morphed Trustees who resembled the original player they previously learned was untrustworthy. Critically, this effect was mediated by the amygdala. The more the morphed picture resembled the untrustworthy player, the more the amygdala activation increased. As FeldmanHall et al. (2018) indicate, in theory, such a learning mechanism might facilitate potentially adaptive decisions to distrust unfamiliar others on the basis of their facial features.

The sketched learning account might also help to explain why adolescents did not yet show the increased amygdala activations when facing trust abusers. As De Neys et al. (2015) noted, facial features are highly unstable during adolescence: The geometry of the human face does not stabilize until age 16 or later (Bulygina, Mitteroecker, & Aiello, 2006; Coquerelle et al., 2011). Since in Western cultures people typically interact with same-age peers (Konner, 2010), adolescents in our study (which was run in



France) will by definition have a harder time to track correlations between facial features and behavior. Their face-trait mapping will be noisier (Over & Cook, 2018). As a result, their amygdala response when faced with untrustworthy adults will be less reliable.

This also points to an interesting possible extension of the present work. In the current study both adolescents and adults played the Trust Game with *adult* Trustees. Adolescents' might not have yet built up their representations of adult facial features in 'trait space' due to lack of experience with adults, but they may have such representations with adolescent faces due to greater contact with same-age adolescent peers. Future studies should therefore also examine adolescents' (and adults) trustworthiness detection when playing with adolescent Trustees.

Taken together, we believe that the present findings lend credence to a possible role of the amygdala in honest signaling of Trustees' trustworthiness on the basis of their facial features. Results suggest that at least in the specific context of the Trust Game, an objective increased risk of being exploited is reflected in an increased amygdala activation among adults.

#### **OPEN DATA STATEMENT**

All behavioral and imaging data can be retrieved from <https://osf.io/q9em6/>

## REFERENCES

- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*(6684), 470-474.
- Adolphs, R., & Tusche, A. (2017). From Faces to Prosocial Behavior: Cues, Tools, and Mechanisms. *Current directions in psychological science*, *26*(3), 282-287.
- Ashburner, J. (2007). A fast diffeomorphic image registration Algorithm. *NeuroImage*, *38*(1), 95-113.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122-142.
- Bohnet, Iris, and Richard Zeckhauser. 2004. Trust, risk and betrayal. *Journal of Economic Behavior and Organization*, *55*, 467-84.
- Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General*, *142*, 143-150.
- Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2015). Face-ism and kernels of truth in facial inferences. *Trends in Cognitive Sciences*, *19*, 421-422.
- Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2017). Can we detect cooperators by looking at their face?. *Current Directions in Psychological Science*.
- Bulygina, E., Mitteroecker, P., & Aiello, L. (2006). Ontogeny of facial dimorphism and patterns of individual development within one human population. *American Journal of Physical Anthropology*, *131*, 432-443.
- Bzdok, D., Langner, R., Caspers, S., Kurth, F., Habel, U., Zilles, K., Laird, A., & Eickhoff, S. B. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure and Function*, *215*(3-4), 209-223.
- Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2015). Honest signaling in trust interactions: smiles rated as genuine induce trust and signal higher earnings opportunities. *Evolution and Human Behavior*, *36*, 8-16.
- Ciriolo, E. (2007). Inequity aversion and trustees' reciprocity in the trust game. *European Journal of Political Economy*, *23*, 1007-1024.
- Coquerelle, M., Bookstein, F. L., Braga, J., Halazonetis, D. J., Weber, G., & Mitteroecker, P. (2011). Sexual dimorphism of the human mandible and its association with dental development. *American Journal of Physical Anthropology*, *145*, 192-202.

- De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2013). Low second-to-fourth digit ratio predicts indiscriminate social suspicion, not improved trustworthiness detection. *Biology Letters*, *9*, 20130037.
- De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2015). Adolescents gradually improve at detecting trustworthiness from the facial features of unknown adults. *Journal of Economic Psychology*, *47*, 17-22.
- De Neys, W., Hopfensitz, A., & Bonnefon, J. F. (2017). Split-second trustworthiness detection from faces in an economic game. *Experimental Psychology*, *64*, 231-239.
- Efferson, C., & Vogt, S. (2013). Viewing men's faces does not lead to accurate predictions of trustworthiness. *Scientific Reports*, *3*, 1-7.
- 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*, 1508-1519.
- FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences*. Advance online publication.
- Fett, A. K. J., Gromann, P. M., Giampietro, V., Shergill, S. S., & Krabbendam, L. (2012). Default distrust? An fMRI investigation of the neural development of trust and cooperation. *Social cognitive and affective neuroscience*, *9*(4), 395-402.
- Freeman, J. B., Stolier, R. M., Ingbreetsen, Z. A., & Hehman, E. (2014). Amygdala responsivity to high-level social information from unseen faces. *Journal of Neuroscience*, *34*, 10573-10581.
- Friston, K., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., & Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *3*(3), 165-189.
- Gee, D. G., Humphreys, K. L., Flannery, J., Goff, B., Telzer, E. H., Shapiro, M., Hare, T. A., Bookheimer, S. Y., & Tottenham, N. (2013). A developmental shift from positive to negative connectivity in human amygdala–prefrontal circuitry. *Journal of Neuroscience*, *33*(10), 4584-4593.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, *115*, 811-846.
- Johnson, N., & Mislin, A. (2011). Trust games: a meta-analysis. *Journal of Economic Psychology*, *32*, 865-889.
- Konner, M. (2010). *The evolution of childhood: Relationships, emotion, mind*. Cambridge, MA: Harvard University Press.

- Little, A. C., Jones, B. C., DeBruine, L. M., Dunbar, I. M. (2013). Accuracy in discrimination of self-reported cooperators using static facial information. *Personality and Individual Differences, 54*, 507-512.
- McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition, 107(1)*, 343-352.
- Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2012). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience, 8(3)*, 285-299.
- Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron, 75(1)*, 73-79.
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7(6)*, 615-631.
- Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. *Trends in Cognitive Sciences, 18*, 566-570.
- Over, H., & Cook, R. (2018). Where do spontaneous first impressions of faces come from?. *Cognition, 170*, 190-200.
- Pessoa, L., & Adolphs, R. (2011). Emotion and the brain: multiple roads are better than one. *Nature Reviews Neuroscience, 12(7)*, 425.
- Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological science, 19(2)*, 109-111.
- Rule, N. O., Krendl, A. C. Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104*, 409-426.
- Rule, N. O., Moran, J. M., Freeman, J. B., Whitfield-Gabrieli, S., Gabrieli, J. D., & Ambady, N. (2011). Face value: Amygdala response reflects the validity of first impressions. *NeuroImage, 54(1)*, 734-741.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9-34.
- Santos, S., Almeida, I., Oliveiros, B., & Castelo-Branco, M. (2016). The role of the amygdala in facial trustworthiness processing: A systematic review and meta-analyses of fMRI studies. *PLoS one, 11(11)*, e0167276.
- Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*, 349-354.

- Thielmann, I., & Hilbig, B. E. (2015). Trust: An integrative review from a person–situation perspective. *Review of General Psychology, 19*, 249.
- Todorov, A. (2017). *Face value: The irresistible influence of first impressions*. Princeton University Press.
- Todorov, A., Funk, F., & Olivola, C. Y. (2015). Response to Bonnefon et al.: Limited ‘kernels of truth’ in facial inferences. *Trends in Cognitive Sciences, 19*, 422-423.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*, 1623-1626.
- Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current opinion in neurobiology, 23(3)*, 373-380.
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2014). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*, 519-545.
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 27*, 813-833.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*, 455-460.
- Tognetti, A., Berticat, C., Raymond, M., & Faurie, C. (2013). Is cooperativeness readable in static facial features? An inter-cultural approach. *Evolution and Human Behavior, 34*, 427-432.
- Tomasello, M. (2009). *Why We Cooperate*. Cambridge, MA: MIT Press.
- Tottenham, N., & Gabard-Durnam, L. J. (2017). The developing amygdala: a student of the world and a teacher of the cortex. *Current opinion in psychology, 17*, 55-60.
- van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A., & Crone, E. A. (2011). Changing brains, changing perspectives: the neurocognitive development of reciprocity. *Psychological Science, 22(1)*, 60-70.
- Van Lange, P. A. M. (2015). Generalized trust: Four lessons from genetics and culture. *Current Directions in Psychological Science, 24*, 71-76.
- van 't Wout, M., & Sanfey, A. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*, 796-803.
- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human brain mapping, 30(3)*, 829-858.

- Villain, N., Landeau, B., Groussard, M., Mevel, K., Fouquet, M., Dayan, J., Eustache, F., Desgranges, B., & Chételat, G. (2010). A simple way to improve anatomical mapping of functional brain imaging. *Journal of NeuroImaging, 20*, 324-333.
- Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage, 18*(2), 293-309.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. *Psychological Science, 17*, 592-598.
- Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal sentencing outcomes. *Psychological Science, 26*, 1325-1331.
- 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.
- Zebrowitz, L. A., Ward, N., Boshyan, J., Gutchess, A., & Hadjikhani, N. (2018). Older adults' neural activation in the reward circuit is sensitive to face trustworthiness. *Cognitive, Affective, & Behavioral Neuroscience*. Advance online publication.

**Table 1**

**Behavioral results.** Proportion (SD) of transfer to cooperators, abusers, and neutral Trustees in both age groups.

Trustee Strategy	Age	
	Adolescents	Adults
Cooperator	52.94 (11.29)	48.27 (17.39)
Abuser	49.05 (21.54)	48.07 (22.58)
Neutral	46.19 (13.71)	45.56 (14.55)

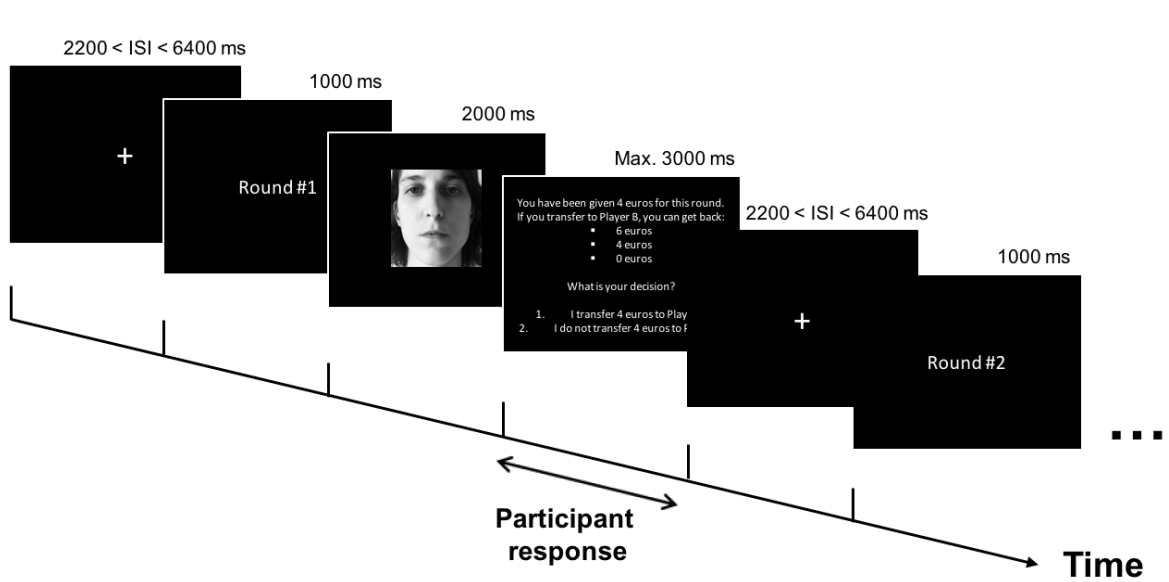
**Table 2**

**Whole-brain results (Adults > Adolescents) for the contrasts of interest.** t-values refer to significant activation peaks at  $p < .001$  (uncorrected for multiple comparisons). All reported clusters were significantly active at  $p < .05$  (FWE). Only regions larger than 4 voxels were considered.

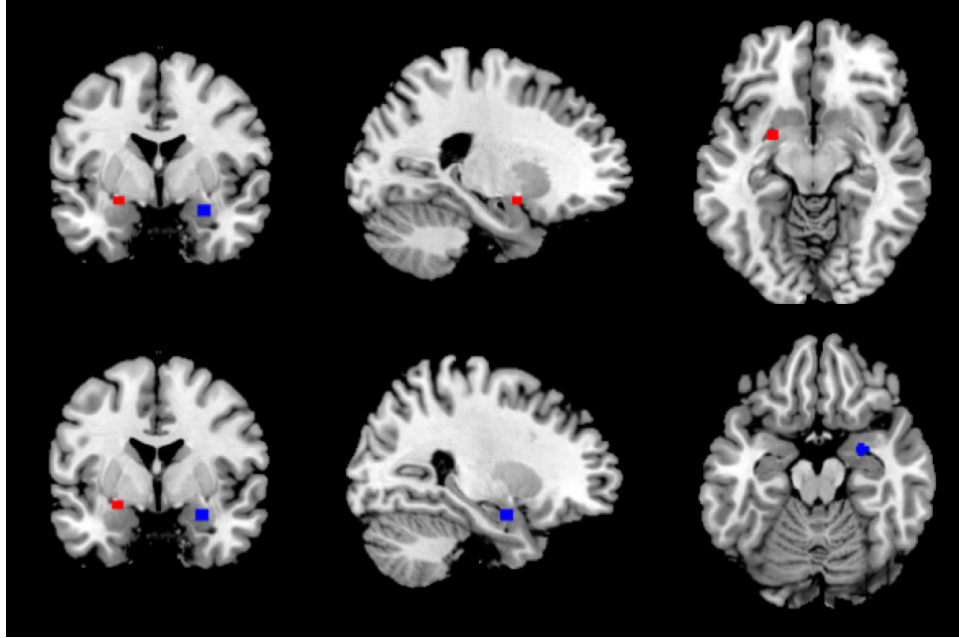
Contrast	Region	Side	X (mm)	Y (mm)	Z (mm)	k	t-value
<b>Abusers &gt; Cooperators</b>	ITG	L	-60	-6	-18	230	5.44
	MFG	L	-30	18	52	1754	5.44
	Angular gyrus	L	-46	-70	34	753	4.89
	Amygdala	L	-28	4	-20	57	4.13
<b>Abusers &gt; Neutral</b>	Frontal lobe (superior part)	L	-22	-10	60	267	5.52
	Thalamus	L	-14	-16	10	311	5.30
	SMA	L	-4	22	44	557	5.24
	Insula	L	-34	-2	-8	902	4.84
	MFG	L	-44	24	36	512	4.82
	IPL	L	-38	-54	56	585	4.59
	Amygdala	L	-26	2	-18	73	4.52
	Precentral gyrus	R	32	-2	50	315	4.44
	Cingulate gyrus (middle part)	L	-6	2	36	246	4.30
	Precuneus	L	-4	-64	44	210	4.12
<b>Cooperators &gt; Neutral</b>	No activations						

*Note.* SMA: Supplementary Motor Area; ITG: Inferior Temporal Gyrus; MOG: Middle Occipital Gyrus; MFG: Middle Frontal Gyrus; IPL: Inferior Parietal Lobule; L: Left; R: Right. k represents the cluster size.

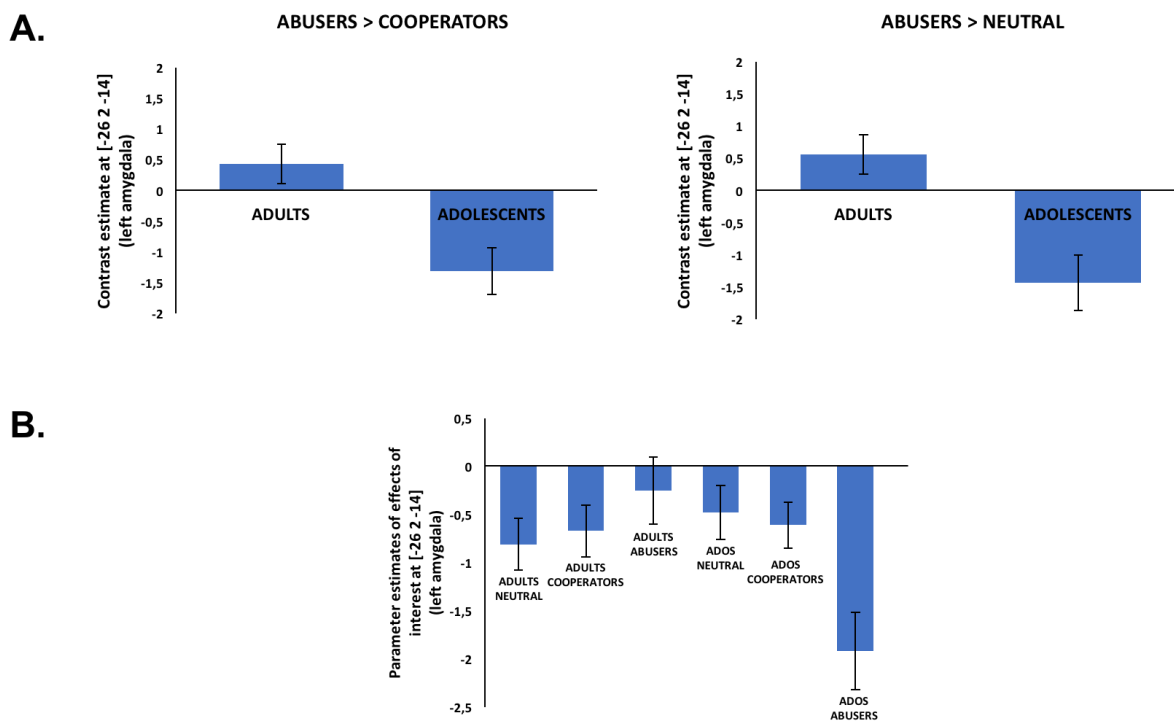




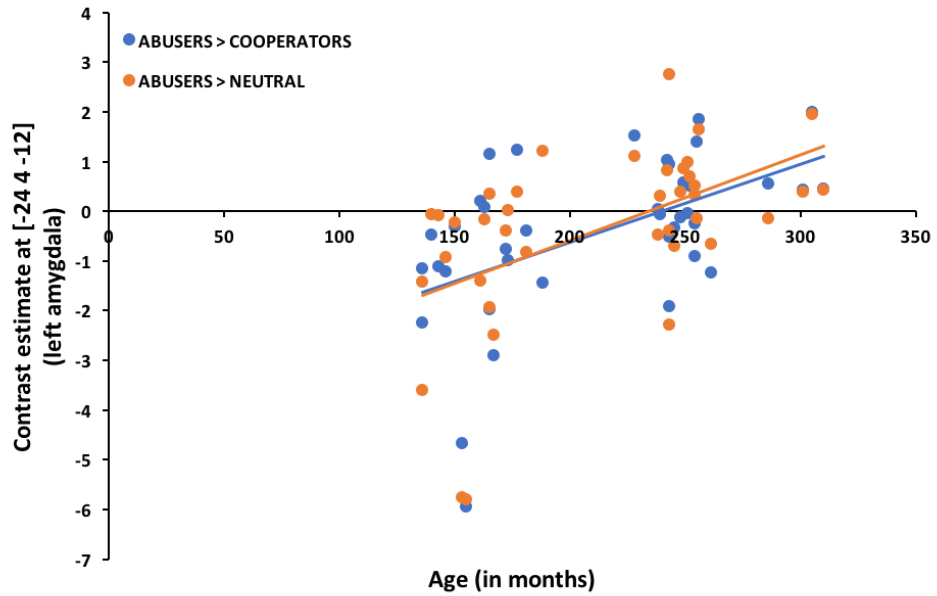
**Figure 1.** Illustration of the Trust Game trial sequence. ISI = Inter Stimulus Interval. Picture is a mock picture presented for illustrative purposes.



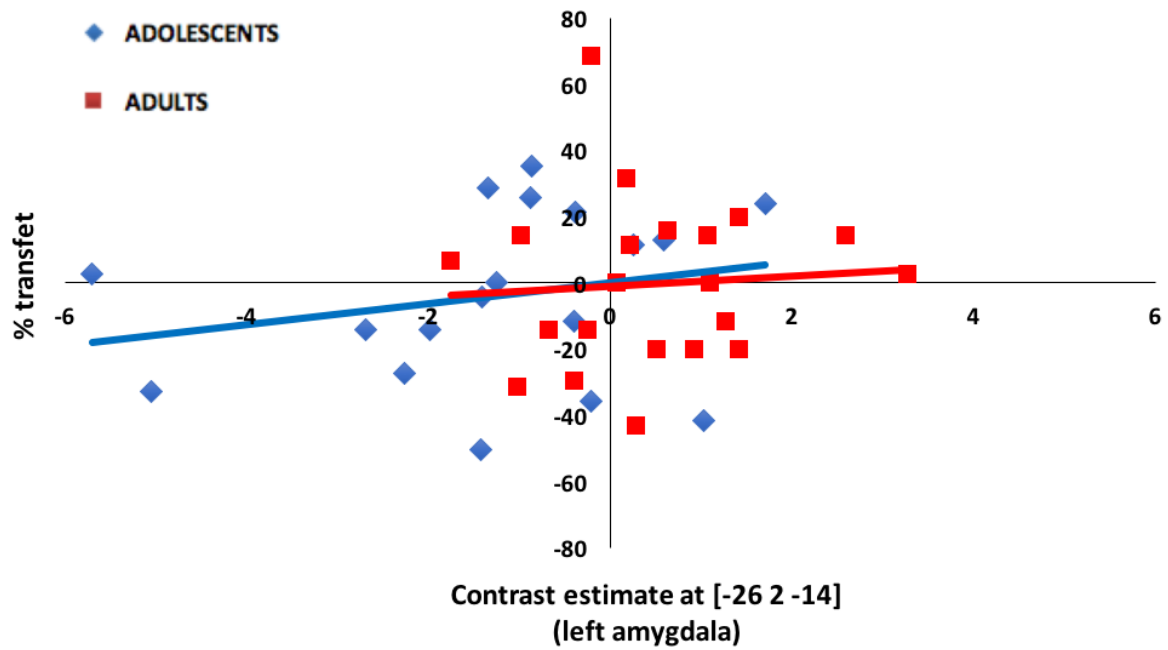
**Figure 2. Amygdala regions of interest (ROIs).** Bilateral amygdala ROIs based on Rule et al. (2013). The ROIs are 4mm radius spheres centered on the following voxels: left amygdala [-24, 0, -12], right amygdala [27, -3, -18]. The left amygdala ROI is depicted in red, the right one in blue.



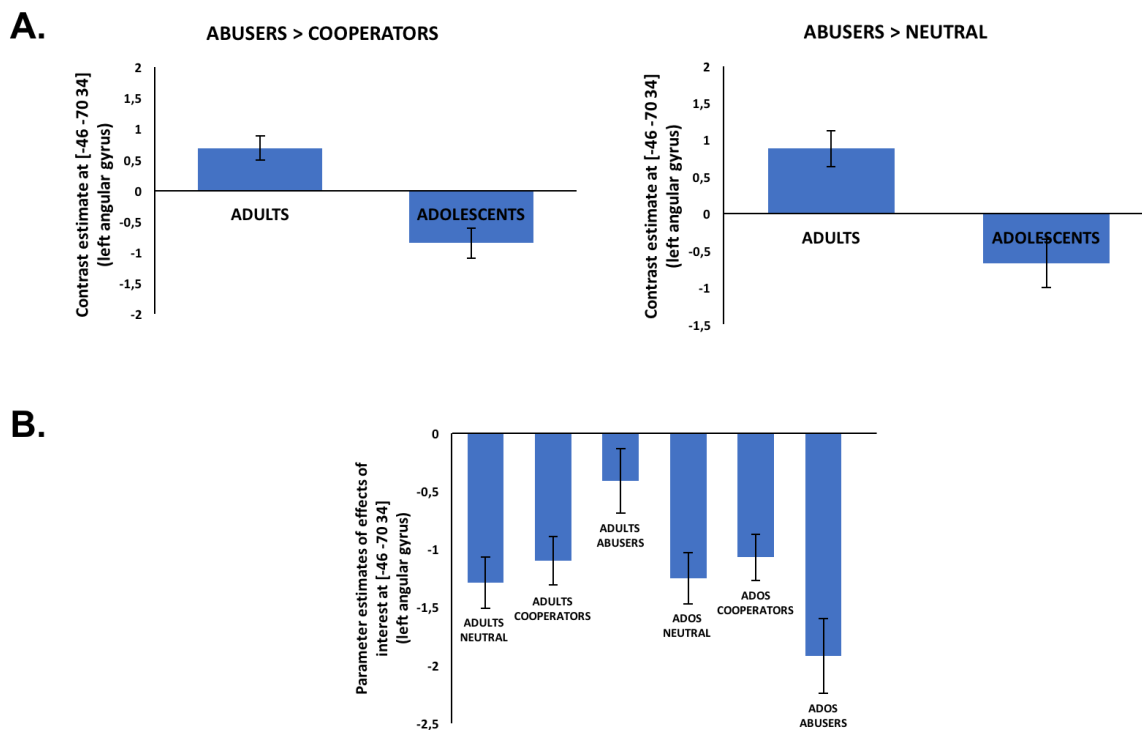
**Figure 3. A. Abusers > Cooperators and Abusers > Neutral contrasts results, in the left amygdala, for adults and adolescents.** Adults show a greater amygdala activation increase when facing a trust Abuser (vs Cooperator or Neutral player) than adolescents. **B. Parameter estimates in the left amygdala for all effects of interest (Neutral, Cooperators and Abusers faces for both age groups).** Adults show less deactivation when looking at trust abusers than at cooperators and neutral players whereas adolescents show the opposite pattern. The error bars represent the standard error of the mean.



**Figure 4.** Correlation between age and left amygdala activation for the **Abusers > Cooperators** and **Abusers > Neutral** contrasts. Results indicate that the amygdala activation in response to presentation of trust abusing Trustees increases with age.



**Figure 5. Correlation between behavioral performance (% transfer to Abuser - % transfer to Cooperator) and left amygdala activation for the Abusers > Cooperators contrast.** Individuals who show a more pronounced activation of the left amygdala when faced with abusers (vs. cooperators), are not less likely to transfer money to these Trustees at the behavioral level.



**Figure 6. A. Abusers > Cooperators and Abusers > Neutral contrasts results, in the left angular gyrus for adults and adolescents.** Adults show a greater activation increase when facing a trust Abuser (vs Cooperator or Neutral player) than adolescents. **B. Parameter estimates in the left angular gyrus for all effects of interest (Neutral, Cooperators and Abusers faces for both age groups).** Adults show less deactivation when looking at trust abusers than at cooperators and neutral players whereas adolescents show the opposite pattern. Note that this pattern is virtually identical to the response pattern in the left amygdala (see Figure 3). The error bars represent the standard error of the mean.