

Social and non-social relational reasoning in adolescence and adulthood

Lucía Magis-Weinberg¹, Sarah-Jayne Blakemore², Iroise Dumontheil³

¹Department of Experimental Psychology, University College London

²Institute of Cognitive Neuroscience, University College London

³Department of Psychological Sciences, Birkbeck, University of London

Corresponding author:

Dr. Iroise Dumontheil

Department of Psychological Sciences

Birkbeck, University of London

Malet Street

London WC1E 7HX

i.dumontheil@bbk.ac.uk

0044 2030738008

Abstract

Reasoning during social interactions requires the individual manipulation of mental representations of one's own traits and those of other people, as well as their joint consideration (relational integration). Research using non-social paradigms has linked relational integration to activity in the rostralateral prefrontal cortex (RLPFC). Here, we investigated whether social reasoning is supported by the same general system or whether it additionally relies on regions of the social brain network, such as the medial prefrontal cortex (MPFC). We further assessed the development of social reasoning. In the social task, participants evaluated themselves or a friend, or compared themselves with their friend, on a series of traits. In the non-social task, participants evaluated their hometown or another town, or compared the two. In a behavioural study involving 325 participants (11-39 years), we found that integrating relations compared to performing single relational judgements improves during adolescence, both for social and non-social information. Thirty-nine female participants (10-31 years) took part in a neuroimaging study using a similar task. Activation of the relational integration network, including the RLPFC, was observed in the comparison condition of both the social and non-social tasks, while MPFC showed greater activation when participants processed social as opposed to non-social information across conditions. Developmentally, the right anterior insula showed greater activity in adolescents compared with adults during the comparison of non-social vs. social information. This study shows parallel recruitment of the social brain and the relational reasoning network during the relational integration of social information in adolescence and adulthood.

Keywords

Relational reasoning, social cognition, development, anterior prefrontal cortex, adolescence

Introduction

Is London more expensive than Cambridge? Answering this question entails at least two levels of relational reasoning. At the first level, one needs to judge the prices in each city independently (*evaluation of single relations, e.g. how much do houses in London cost?*). At the second level, one needs to simultaneously consider mental representations of both cities, and to integrate the single judgements into a higher-order comparison (*relational integration, in this case, comparing the house prices in London and Cambridge*). Relational integration has typically been studied in non-social contexts, in particular using the Raven's Progressive Matrices (Raven, 1998). However, relational integration also occurs in the social domain, for example, when comparing people on personality traits (*e.g. are you more patient than your friend?*). The neural processes supporting this kind of social reasoning, and the way it develops, are not well understood.

Previous functional magnetic resonance imaging (fMRI) research has identified lateral prefrontal cortex (PFC) and lateral parietal cortex as involved in relational integration (Bunge, Helskog, & Wendelken, 2009; Dumontheil, 2014; Smith, Keramatian & Christoff, 2007; Wendelken, Nakhabenko, Donohue, Carter & Bunge, 2008), while medial PFC has been associated with the processing and manipulation of social information (Gilbert et al., 2006; Van Overwalle, 2009; Wood & Grafman, 2003). The current study aimed to bring together these separate strands of research to investigate domain-general and social domain-specific processes that support the relational integration of social information. Both relational reasoning and social cognition and their underlying neural substrates undergo significant reorganization during adolescence (Dumontheil, 2014; Kilford, Garrett & Blakemore, 2016). Therefore, a second aim of the current study was to compare social reasoning in adolescents and adults. We employed a paradigm that allows the investigation and comparison of relational integration of both social and non-social information (Raposo, Vicens, Clithero, Dobbins, & Huettel, 2011). In a large behavioural study, we investigated the development of relational integration of social and non-social information from late childhood until adulthood. In

a follow-up fMRI study, we studied the neural correlates of these cognitive processes in adolescence and adulthood.

Neural bases of relational integration and social cognition

Relational reasoning research suggests a central role of rostralateral prefrontal cortex (RLPFC), which corresponds to the lateral aspect of the anterior, or rostral, PFC (Brodmann areas (BA) 10/46 and 10/47), in relational integration compared with processing single relations. Imaging studies using the Raven's Progressive Matrices in adults have shown RLPFC involvement in the joint manipulation of visuospatial patterns (Christoff et al., 2001; Kroger et al., 2002), as well as in the integration of relations in analogical reasoning tasks (Bunge, Wendelken, Badre, & Wagner, 2005; Wendelken, et al., 2008) and in the integration of multiple relations to reach a logical conclusion (Wendelken & Bunge, 2009). A study comparing visuospatial and semantic variants of a relational matching task found considerable activation overlap within left RLPFC, suggesting a domain-general role for RLPFC in relational integration (Wendelken, Chung, & Bunge, 2012).

Social cognitive research suggests a role of the medial prefrontal cortex (MPFC), which corresponds to BA 8/9/10, in the processing of social information (see Van Overwalle, 2009 for a meta-analysis). Studies with adults have shown that this region is involved in considering one's thoughts and feelings (Gusnard, 2005; Rameson, Satpute, & Lieberman, 2010; Zysset, Huber, Samson, Ferstl, & von Cramon, 2003) and in perspective taking (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006; D'Argembeau et al., 2007; David et al., 2008; David et al., 2006; Ruby & Decety, 2001, 2004; Vogeley et al., 2004). The MPFC is also recruited during tasks that require mentalising, that is, the consideration of other people's mental states (Amodio & Frith, 2006; Decety & Sommerville, 2003; Frith & Frith, 2003).

Development during adolescence

Both relational integration and social cognition show protracted development in terms of improved performance and associated brain activity between adolescence and adulthood (Blakemore, 2012; Crone et al., 2009; Crone & Dahl, 2012; Dumontheil & Blakemore, 2012; Dumontheil, Burgess, & Blakemore, 2008, Dumontheil, Hillebrandt, Apperly, & Blakemore, 2012). The RLPFC undergoes structural and functional development with age, with evidence that its activity during relational integration tasks becomes increasingly specialised with age (Crone et al., 2009; Dumontheil, 2014; Dumontheil et al., 2008; Dumontheil, Houlton, Christoff, & Blakemore, 2010; Ferrer, O'Hare, & Bunge, 2009; Wendelken, O'Hare, Whitaker, Ferrer, & Bunge, 2011). In addition, a complex pattern of developmental changes in functional connectivity related to reasoning ability has been identified, including changes in connectivity between the RLPFC and the parietal cortex (Bazargani, Hillebrandt, Christoff, & Dumontheil, 2014; Wendelken, Ferrer, Whitaker, & Bunge, 2015). Bazargani et al. (2014) observed a decrease in short-range (fronto-insular) connectivity with stable long-range connectivity (frontoparietal) and an increase of modulatory connections with age. Wendelken et al. (2015) found a pattern of developmental changes suggestive of increasing communication between prefrontal regions and specific targets.

Key regions of the social brain, including the MPFC, undergo structural and functional changes during adolescence. Cortical thickness and grey matter volume in the MPFC decrease between late childhood and the early twenties (Mills, Lalonde, Clasen, Giedd and Blakemore, 2014). In parallel, several fMRI studies have shown that MPFC activity during mentalising tasks decreases between early adolescence and adulthood (Blakemore, 2008, 2012). In a previous study investigating the development of the neural correlates of mentalising, participants were required either to take someone else's perspective or to use symbolic cues to select an appropriate action in a communicative context. We found that adolescents showed hypoactivation of domain-general cognitive control regions in the parietal cortex and PFC, and hyperactivation of parts of the social brain network (Dumontheil, et al., 2012). This study thus demonstrated the engagement of cognitive

control and social brain regions within a single paradigm, and that the engagement of these regions changes as a function of age.

Relational integration within the social domain has previously been investigated in adults using a task that combined both mentalising and relational integration (Raposo et al., 2011). Participants judged how pleasant they found a certain word, how pleasant a friend would find the word, and how their rating of pleasantness would compare to that of their friend. Behaviourally, reaction times were higher when participants were comparing themselves to their friend relative to the two single relations conditions. MPFC activation was higher during the friend judgement compared with the self judgement, while dorsolateral frontal cortex activation was higher when contrasting the relational integration comparison and the self judgement conditions. The study did not include a non-social relational reasoning condition, preventing the conclusion that the activation patterns are specific to relational integration of social information per se, or reflective of relational integration more generally.

Present study

Here, we adapted the paradigm designed by Raposo et al. (2011) to investigate behavioural development of social reasoning (Study 1) and its neural development between adolescence and adulthood (Study 2). We compared first-order judgements (1-REL) of traits associated with oneself or with another individual (e.g. How patient are you? (Self condition); How patient is your friend? (Other condition)), with second-order judgements (2-REL) about how these judgements related to each other (How much more patient are you than your friend? (Comparison condition)). Our paradigm also included a control non-social task, in which participants were asked to rate characteristics of towns. Our aim was to assess: (1) how performance on a task requiring relational integration of social or non-social traits develops between late childhood and adulthood; (2) how neural activity underlying these processes develops between early adolescence and adulthood; and

(3) whether there is domain-specific activation for the relational integration of social vs non-social information.

In terms of behaviour, we predicted improvements in relational integration with age, both in terms of reaction time and the consistency of participants' responses between 1-REL and 2-REL judgements. In terms of blood-oxygen level dependent (BOLD) signal, we expected domain-general activations associated with relational integration in RLPFC, dorsolateral PFC and parietal cortex. We also predicted there would be additional domain-specific activations in parts of the social brain network associated with the People task, specifically regions involved in processing social information and mentalising (Dumontheil & Blakemore, 2012; Meyer, Spunt, Berkman, Taylor, & Lieberman, 2012; Meyer, Taylor, Lieberman, 2015; Raposo et al., 2011). Finally, we predicted that the RLPFC would show increased specificity of activation for REL-2 vs. REL-1 judgements in adults compared to adolescents (Dumontheil, 2014), and that MPFC would show greater activation in adolescents than adults in the social vs. non-social task (Blakemore, 2008; Blakemore & Robbins, 2012).

Study 1 – Behavioural Study

Participants

The data analysed here are part of a larger project in male and female children, adolescents and adults who performed a set of six tasks and provided saliva samples for genetic analyses (Dumontheil et al., 2014; Kilford, Dumontheil, Wood, & Blakemore, 2015). The present analysis focused on the social and non-social comparison task and the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) assessment. The data presented here are from 325 participants aged between 11 and 39 years old (N = 160 adults, N = 165 children and adolescents). From an original sample of 399 participants, one was excluded because of a diagnosis of Turner Syndrome, one was excluded because of a diagnosis of Asperger Syndrome, four were excluded because of a task

programming error, two were excluded because they interrupted the task early, one participant did not have time to complete this task and data were lost from 15 participants. In addition, as only four male participants were younger than 11 years old (vs. 26 female participants) all participants younger than 11 years old were excluded from further analyses. Children and adolescents were recruited from schools in and around London and were tested in their school, and adults were recruited from the University College London (UCL) Psychology Department volunteer database and word of mouth and tested in the lab. Written informed consent was obtained from participants or from the parent/guardian of participants under 18. Adult participants were remunerated for their time. The study was approved by the UCL Ethics Committee.

Child and adolescent participants were divided according to age into three groups spanning two or three years, and there were four adult groups (**Table 1**). Verbal ability was measured using the Vocabulary subtest of the WASI (Wechsler, 1999). A two-way (Age group, Gender) ANOVA indicated there was a significant difference in verbal IQ between age groups ($F(1, 307) = 2.91, p = .009$). Paired post-hoc comparisons demonstrated that the 11-12y and 13-14y groups had lower mean verbal IQ than the 20-22y group ($p < .05$) and that the 26-28y group had lower mean verbal IQ than all other age groups ($p < .05$) (**Table 1**). Only the difference between 20-22y and 26-28y groups survived Bonferroni correction ($p = .001$). There was no main effect of Gender on IQ ($F(1, 307) = 2.34, p = .127$), but a significant Gender x Age group interaction ($F(6, 307) = 3.08, p = .006$). Post-hoc comparisons of males and females in each age group indicated that 11-12y males had lower mean verbal IQ (109.1, $SD = 12.2$) than females (118.1, $SD = 10.5, p = .002$), while 23-25y males had higher mean verbal IQ (118.6, $SD = 11.1$) than females (109.7, $SD = 14.5, p = .023$).

Insert Table 1 here

Design and Stimulus Material

The task had two within-subjects factors (Task: People or Town; Condition: Self, Other or Comparison) and one between-subjects factor (Age group: seven levels), resulting in a 2 x 3 x 7 mixed factorial design. The task was computer-based and adapted from the fMRI study by Raposo et al. (2011). The task was administered as part of a single individual testing session of approximately 45-50 minutes. It was the third task administered in the task set, and the WASI was administered as the sixth (and last) task.

The experimenter started by asking participants to think of someone that they knew quite well but who was quite different from them, and to give his/her name. If participants did not respond, the experimenter suggested they consider a close friend or a sibling who was quite different from them. Second, participants were asked to name the town where they lived (typically London) and then pick a town that they knew quite well but which was quite different from London. Again, if participants did not respond, the experimenter suggested they consider a town where they go on holiday or where their grandparents live. Instructions were then presented on the screen and read aloud to the participants, explaining the different types of judgement they would make during the task, and the rating scale. Participants used the index, middle and ring fingers of both hands to respond. The task was programmed in Cogent (http://www.vislab.ucl.ac.uk/cogent_graphics.php) running in MatLab (MathWorks) on a DELL 12 inch laptop or similar.

Judgements were blocked according to Task and Condition and the order of the blocks was counterbalanced within and between participants. Each block started with an instruction screen indicating to participants what type of judgement they should make during that block. On each trial, this information was repeated at the top (e.g. "You"), an adjective was presented in the middle of the screen, and a rating scale from 1 to 6 was provided at the bottom of the scale (**Figure 1**).

Insert Figure 1 here

Once participants had pressed a key to indicate their response, the corresponding number on the scale (1-6) was highlighted in red for 200 ms, followed by a 200 ms blank screen, and then a new trial started. There were 10 trials per block and 12 blocks in total, with two blocks of each of the six conditions (People, Town x Self, Other, Comparison), i.e. 20 trials in total per condition. A list of 20 adjectives was used for this study (fabulous, weird, loud, charming, romantic, crazy, pleasant, lovely, wild, perfect, busy, unique, friendly, cool, unusual, boring, dull, rich, quiet, popular). All adjectives were presented once in each condition of each task.

Data Analysis

Data were analysed with SPSS 21 (Armonk, NY: IBM Corp.). Median RTs were calculated for each Task and Condition. A 2 (Task) x 3 (Condition) x 7 (Age group) x 2 (Gender) mixed-model, repeated measures ANOVA (rmANOVA) was performed on the median RT data. Due to the nature of the task, a measure for accuracy is not available (we do not know objectively how “friendly” a participant or their friend is). Instead, we ran Spearman correlations to test whether participants were consistent in their response, i.e. whether there was a positive correlation between their ratings in the Comparison judgement and the difference in ratings between the Self and Other judgements, across the set of 20 adjectives. For example, if a participant rated herself as “6” and her friend as “3” for “friendly”, a rating of “5” in the Comparison judgement (“How much more friendly do you think you are compared to your friend?”) would be consistent with the individual judgements. Spearman r correlation values obtained for each participant in each task were transformed using Fisher’s z transformation ($z = 0.5 \times \ln \left(\frac{1+r}{1-r} \right)$), resulting in consistency scores. A 2 (Task) x 7 (Age group) x 2 (Gender) mixed-model rmANOVA was performed on the consistency data. Post-hoc comparisons investigating pairwise differences between age groups were Bonferroni corrected. Greenhouse-Geisser correction for non-sphericity was applied when necessary. Estimated means and standard errors from the rmANOVAs are reported in the text and in figures.

Results

Reaction times

A 2 (Task) x 3 (Condition) x 7 (Age group) x 2 (Gender) mixed rmANOVA showed a main effect of Condition ($F(1.4, 435.6) = 500.00, p < .001, \eta_p^2 = .62$), whereby participants responded faster in the Self ($M = 1943 \pm 36$ ms) than the Other ($M = 1989 \pm 35$ ms) condition ($p = .042$) and in both Self and Other conditions than in the Comparison condition ($M = 2811 \pm 56$ ms; $ps < .001$), a main effect of Task ($F(1, 311) = 12.04, p < .001, \eta_p^2 = .04$), with slower RTs in the People ($M = 2292 \pm 44$ ms) than the Town task ($M = 2203 \pm 39$ ms), and a main effect of Age Group ($F(6, 311) = 2.48, p = .024, \eta_p^2 = .05$), but no main effect of Gender ($F(1, 311) = .053, p = .818$). Pairwise post-hoc comparisons indicated that the 11-12y group responded more slowly than the 20-22y group ($p = .019$) and marginally slower than the 26-28y group ($p = .081$). No other pairwise comparison was significant.

The only significant interaction was between Condition and Age group ($F(8.4, 435.6) = 3.64, p < .001, \eta_p^2 = .07$). This was followed up by first entering the two 1-REL judgements in a 2 (Task) x 7 (Age group) x 2 (Gender) rmANOVA. This analysis showed no main effect of Age group ($F(6, 311) = 1.54, p = .165$) and no Age group x Condition interaction ($F(6, 311) = 1.35, p = .233$), indicating that the interaction was driven by the 2-REL, Comparison condition. Investigated separately, this condition showed a main effect of Age group ($F(6, 311) = 3.61, p = .002, \eta_p^2 = .07$). Post-hoc comparisons indicated that the 11-12y group responded slower on average than the 20-22y and 23-25y groups ($ps < .01$) (**Figure 2A**). To further investigate the interaction, the difference in median RT between the Comparison condition (2-REL, relational integration) and Self and Other conditions (1-REL judgements) was calculated. A 2 (Task) x 7 (Age group) x 2 (Gender) rmANOVA with post-hoc pairwise comparisons indicated that the 11-12y group was relatively slower in 2-REL than 1-REL trials than the 13-14y, 20-22y, 23-25y and 26-28y groups ($ps < .05$) (**Figure 2B**).

Insert Figure 2 here

Consistency

A 2 (Task) x 7 (Age group) x 2 (Gender) mixed rmANOVA was performed on consistency, which is the Fisher z-transformed Spearman correlation between participants' answers in the Comparison condition and the difference between their answers in the Self and Other conditions. The main effect of Task was significant ($F(1, 311) = 21.74, p < .001, \eta_p^2 = .07$), with greater consistency in the Town ($M = .713 \pm .025$) than in the People task ($M = .586 \pm .021$). In both conditions, mean consistency was positive, indicating some degree of consistency in participants' responses across conditions. The main effect of Age group was also significant ($F(6, 311) = 15.15, p < .001, \eta_p^2 = .23$). No other main effect and no interaction was significant. Post-hoc comparisons indicated that the 11-12y and 13-14y groups did not differ from each other and were less consistent than all older age groups ($ps < .05$), which did not differ from each other (**Figure 2C**).

In summary, participants were slightly slower in the Other than in the Self condition, and in the People than in the Town task. However, there was a considerable difference in RT between 2-REL and 1-REL judgements, which decreased between 11-12 and 13-14 years of age. Consistency was also higher in the Town than in the People task, and improved between 13-14 and 15-17 years. There was no interaction between Task and Age group, suggesting no evidence of a developmental difference in relational integration between the social and non-social tasks.

Study 2 – Neuroimaging Study

Methods

Participants

Thirty-nine female participants aged between 10 and 31 years ($N = 20$ adults, $N = 19$ children and adolescents), took part in the neuroimaging study (**Table 1**). Seven adolescents and two adults had previously taken part in Study 1, with an interval of between 4 and 10 months between the testing sessions. Only female participants were included to reduce variability in the sample due to sex differences in brain development (Herting, Maxwell, Irvine, & Nagel, 2012; Raznahan et al., 2011).

Gender differences were not found in Study 1 or in a previous behavioural and neuroimaging study of visuospatial relational reasoning development (Wendelken et al., 2011). However, other behavioural studies have reported gender differences in mentalising (e.g. Charman, Ruffman & Clements, 2002) and in relational reasoning (Lynn & Irwing, 2004). Since we were unable to collect a sample large enough to investigate gender differences, we chose to maximize the homogeneity of our sample by only including female participants.

Participants were reimbursed £20 and their travel expenses for taking part in the study. The study was approved by the UCL Research Ethics Committee. Participants were divided into two groups, adolescents and adults. Adolescents were combined into a single group because of the sample size, with a focus on investigating the development of the neural correlates of relational reasoning, while Study 1 had focused on developmental changes in performance. The two groups were matched on estimated IQ ($t(36) = 1.02, p = .314$), which was assessed using the Vocabulary and Matrices subtests of the WASI (Wechsler, 1999).

Design and stimulus material

The fMRI task had two within-subjects factors (Task: People, Town; Condition: Self, Other, Comparison, Vowels) and one between-subjects factor (Age group: adults, adolescents), resulting in a 2 x 4 x 2 mixed design. Participants were first trained on the task outside the scanner. After going through the four types of question for each Task, participants performed one block of three trials for each Task and Condition. Participants then performed four scanning runs as well as a structural scan between the second and third task runs. After scanning, participants completed the Interpersonal Reactivity Index (IRI; Davis, 1980), which provides measures of four components of empathy (empathic concern, fantasy, personal distress and perspective taking), and were assessed on the WASI. The IRI was included as it has been used in previous neuroimaging studies as a measure of individual differences in social cognition in everyday life (Meyer, et al., 2012; Raposo et al., 2011). Meyer et al. (2012) found an association between memory load-dependent activity within

mentalising regions and scores on the Perspective Taking scale. We therefore aimed to relate activity within mentalising regions in Study 2 with this everyday life measure of perspective taking.

The paradigm was similar to the task described in Study 1, adapted from Raposo et al. (2011), with three differences. First, the fMRI task included a Vowels condition in which participants were required to count the number of vowels in the adjective presented on the screen (“Control” condition used by Raposo et al., 2011). This condition matched visual and verbal processing and motor execution demands while minimising relational reasoning demands and, as such, was used as a baseline in the fMRI analyses. Regions of the social brain often show high activation at rest or during fixation phases, as part of the default mode network. Using an active baseline allowed the identification of activation of these brain regions, which was important for the People task. In addition, the words presented in the People and Town tasks differed, and using the Vowels condition as a baseline allowed us to control for BOLD signal differences elicited by the presentation of these words. Finally, using a similar baseline to Raposo et al. (2011) facilitates comparison of the results of the two studies. Second, traits were different for the People and Town tasks in an attempt to make them more relevant to each category. Third, a fixed inter-stimulus-interval (ISI) of 100 ms was used.

Participants performed two scanning runs of the People and Town task, alternating in an ABAB or BABA order, counterbalanced across participants. Each run comprised five blocks of each condition (Self, Other, Comparison and Vowels). The order of the conditions was fixed within a run, and counterbalanced across runs and participants. After every Vowels block, there was a fixation block. Task blocks were preceded by a 1 s instruction that specified the condition of the next block (e.g. “You”, “Sam”, “You compared to Sam”, “Vowels”, “London”, “Cambridge”, “London compared to Cambridge”) and comprised three trials each.

Participants had a maximum of 6.05 s to input their response on each trial, during which time the stimulus remained on the screen. When participants responded, the number they pressed turned

red and the stimulus remained on the screen until 6.1s after the onset of the trial presentation. A blank screen was displayed during the ISI. Stimuli consisted of lists of 30 adjectives in each task, which were matched for number of letters, number of vowels, frequency and familiarity (see **Appendix**). Each adjective was presented once in each of the four conditions. Half of the adjectives were presented in the first scanning run of a task, half in the second run.

FMRI acquisition

Multi-slice T2-weighted echo-planar volumes with BOLD contrast (35 axial slices with a voxel resolution of $3 \times 3 \times 3$ mm covering most of the cerebrum; TR = 2.975 s; TE = 50 ms; TA = 2.925 s) were obtained using a 1.5 T MRI scanner (Siemens TIM Avanto, Erlangen, Germany). Functional images were acquired in four scanning runs lasting approximately 8 min 40 s each in which 174 volumes were obtained. The first four volumes of each run were discarded to allow for T1 equilibrium effects. A 3D T1-weighted fast-field echo anatomical image lasting 5 min 30 s was acquired after the first two functional runs for each participant.

Data Analysis

Behavioural data

A 2 (Task) x 4 (Condition) x 2 (Age group) mixed rmANOVA was performed on median RT data. A 2 (Task) x 2 (Age group) mixed rmANOVA was employed to analyse mean consistency, which was calculated in the same way as in Study 1.

MRI data

MRI data were preprocessed and analysed using SPM8 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, <http://www.fil.ion.ucl.ac.uk/spm/>). Images were realigned to the first analysed volume with a second-degree B-spline interpolation to correct for movement during the session. The bias-field corrected structural image was coregistered to the mean, realigned

functional image and segmented on the basis of Montreal Neurological Institute (MNI)-registered International Consortium for Brain Mapping (ICBM) tissue probability maps. Resulting spatial normalisation parameters were applied to the realigned images to obtain normalised functional images with a voxel size of 3 x 3 x 3 mm, which were smoothed with an 8-mm full width at half maximum Gaussian kernel.

Realignment estimates were used to calculate framewise displacement (FD) for each volume, which is a composite, scalar measure of head motion across the six realignment estimates (Siegel et al., 2013). Volumes with an FD > 0.9 mm were censored and excluded from general linear model (GLM) estimation by including a regressor of no interest for each censored volume. Scanning sessions with more than 10 % of volumes censored or a root mean square (RMS) movement over the whole session greater than 1.5 mm (1 session for three participants, 2 sessions for one participant) were excluded from the analysis. Adolescent and adult participants significantly differed in the number of overall censored volumes ($M_{\text{adolescents}} = 3.39 \pm 3.61$ (SD), $M_{\text{adults}} = 0.19 \pm 0.38$; $p < .001$), mean RMS translational movement ($M_{\text{adolescents}} = 0.32 \text{ mm} \pm 0.11$, $M_{\text{adults}} = 0.24 \text{ mm} \pm 0.07$; $p = .005$), and mean FD ($M_{\text{adolescents}} = 0.18 \text{ mm} \pm 0.08$, $M_{\text{adults}} = 0.10 \text{ mm} \pm 0.02$; $p < .001$). There was no difference between groups in terms of mean RMS rotational movement ($M_{\text{adolescents}} = 0.23 \text{ mm} \pm 0.12$, $M_{\text{adults}} = 0.17 \text{ mm} \pm 0.08$; $p = .088$).

Scanning runs were treated as separate time series and each series was modelled by a set of regressors in the GLM. Runs of the People or Town Task were each modelled by six box-car regressors: four regressors corresponding to each Condition (Self, Other, Comparison, Vowels), with a duration of 18.6 s; Instructions, with a duration of 1 s; and Fixation blocks, with a duration of 18.6 s except for the last block which had a duration of 39 s. All regressors were convolved with a canonical haemodynamic response function and, together with the separate regressors representing each censored volume and the mean over scans, comprised the full model for each session. The data and model were high-pass filtered to a cut-off of 1/128 Hz.

The second-level whole-brain analysis focused on relational integration, i.e. the main effect of Comparison (2-REL) versus Self and Other (1-REL) conditions, and on differences between social and non-social tasks. The 1-REL conditions, Self and Other, were thus combined within each task. Four first level contrasts were calculated using the Vowels condition as a baseline within each task: People [Self, Other] – People Vowels (*People SO*), People Comparison – People Vowels (*People Comp*), Town [Self, Other] – Town Vowels (*Town SO*), Town Comparison – Town Vowels (*Town Comp*). These contrasts were then entered into a random-effects analysis using a Subject x Age group (2) x Block type (4) flexible factorial design, modelling Subject as a main effect (to account for the repeated-measure nature of the data) and the Age group x Block type interaction.

Main effects of Condition (Comparison > SO) and Task (People > Town and Town > People) and the interaction between the two factors and with Age group were determined using the *t* statistic on a voxel-by-voxel basis. Statistical contrasts were used to create SPM maps thresholded at $p < .001$ at the voxel level and at family-wise error (FWE)-corrected $p < .05$ at the cluster-level (corresponding to a minimum cluster size of 77 voxels determined with SPM8). Activations that survived whole-brain FWE correction at $p < .05$ at the voxel-level are indicated. All coordinates are given in MNI space. Significant interactions were followed up by extracting the mean signal across all voxels of significant clusters with MarsBar (Brett, Anton, Valabregue, & Poline, 2002) and analysing simple effects in SPSS using t-tests (with Bonferroni correction for multiple comparisons).

We performed exploratory correlation analyses between consistency in ratings and individual differences in activation in the tasks. Relevant task contrasts were entered in a two sample t-test design modelling the two age groups separately, with the behavioural measure as a single covariate of interest. Correlations were run between the contrasts [Comparison > SO] and [Comparison/Self/Other > Vowels] and the mean consistency across tasks; between the contrast [People Comparison > SO] and consistency in the People task; and between the contrast [Town Comparison > SO] and consistency in the Town task. In the same manner, we explored correlations

between individual differences in the [People > Town] and [People Comparison > SO] contrasts and the perspective-taking scale of the IRI (see Meyer, Taylor, & Lieberman, 2015, for a similar approach).

Results

Behavioural results

Reaction times

There was a main effect of Task ($F(1, 37) = 13.51, p = .001, \eta_p^2 = .27$). In contrast to Study 1, participants were slower in the Town ($M = 2419 \pm 69$ ms (SE)) relative to the People task ($M = 2282 \pm 63$ ms). There was a main effect of Condition ($F(1.8, 66.8) = 29.17, p < .001, \eta_p^2 = .44$) (**Table 2**). Pairwise comparisons with Bonferroni correction revealed that participants were slowest in the Comparison condition relative to all other conditions ($M_{Comparison} = 2638 \pm 79$ ms, $M_{Self} = 2180 \pm 67$ ms, $M_{Other} = 2195 \pm 62$ ms, $M_{Vowels} = 2388 \pm 77$ ms, all p 's < .005). The Vowels condition was the next slowest (all p 's < .05). Self and Other conditions did not differ significantly from one another ($p > .05$). In contrast to Study 1, there was no main effect of Age group ($F(1,37) = 2.21$).

Insert Table 2 here

There was a significant interaction between Task and Condition ($F(3, 111) = 3.14, p = .028, \eta_p^2 = .08$). This was followed up by analysing the data in the People and Town tasks separately. In both the People and Town tasks, participants were slower in Comparison than Self and Other trials (all p 's < .001). In the People task, participants were slower in Vowels than Self and Other trials (all p 's < .05). The two-way interaction is driven by a greater difference between Comparison and Vowels trials in the Town ($M_{Comp-Vowels} = 342$ ms) than in the People task ($M_{Comp-Vowels} = 157$ ms, $p = .008$) while the difference between Comparison and Self and Other trials did not differ between tasks (p 's > .05). In contrast to Study 1, there was no significant interactions between Condition and Age group ($F(1.8, 66.8) = 0.55$). In line with Study 1, there were no significant interactions between Task and Age

group ($F(1,37) = 1.43, p = .24, \eta_p^2 = .04$) or amongst Condition, Task and Age group ($F(2.7, 98.8) = 1.76$).

Consistency

As in Study 1, participants were consistent overall, with positive correlations between the 1-REL and 2-REL ratings. Similar to the pattern in Study 1, there was a trend effect of Age group ($F(1, 37) = 3.46, p = .071, \eta_p^2 = .09$, **Figure 2D**), with lower consistency in adolescents than adults. There was also a main effect of Task ($F(1, 37) = 18.04, p < .001, \eta_p^2 = .33$): consistency was lower for People than for Town ($M_{People} = .74 \pm .05, M_{Town} = .96 \pm .07$). As in Study 1, the interaction between Task and Age group was not significant ($F(1,37) = 0.67$).

IRI

Analyses of the subscales of the IRI revealed significantly higher Perspective Taking (PT) and Empathic Concern (EC) scores for adults ($M_{PT} = 19.5 \pm .91$ (*SE*); $M_{EC} = 20.9 \pm 1.1$) than for adolescents ($M_{PT} = 14.1 \pm .94$; $M_{EC} = 16.8 \pm 1.1$; all p 's < .05). Pearson correlation coefficients were computed to assess the relationship between PT and performance in the task. There was no correlation between PT and Consistency in the People task ($r = -.187, p = .255$). There was also no correlation between PT and mean median RT for the Self ($r = -.021, p = .897$), Other ($r = .006, p = .973$), or Comparison ($r = .068, p = .680$) conditions in the People task.

FMRI Results

Whole-brain analyses contrasted the Comparison condition to the combined Self and Other (SO) conditions in the People and Town tasks. The Vowels condition served as an active baseline.

Relational integration

A broad bilateral network of fronto-parietal, temporal and occipital regions, including bilateral RLPFC, showed increased BOLD signal in Comparison vs SO conditions, (**Table 3** and **Figure 3A**), that

is, in 2-REL as opposed to 1-REL processing. Increases in BOLD signal were observed in a large posterior cluster extending into bilateral occipital and lingual gyri, calcarine sulcus, inferior parietal lobule, and in anterior clusters in the precentral gyrus, pre-supplementary motor area (pre-SMA), and inferior and middle frontal gyri.

Insert Table 3 here

Social information processing

When comparing the People task to the Town task, that is, when the information to be processed was of social (traits of participant or a friend) versus non-social nature (characteristics of towns), increased BOLD signal was observed in clusters in the MPFC, insula, and precuneus (**Table 3** and **Figure 3B**). The reverse contrast revealed increased BOLD signal in a large cluster extending into bilateral calcarine gyri, middle and superior occipital gyri, as well as bilateral clusters in the fusiform and parahippocampal gyri extending into the medial temporal gyri, in the middle cingulate cortex, bilateral pre- and postcentral gyri and left inferior frontal cortex (**Table 3**).

Insert Figure 3 here

Age group differences

There was no two-way interaction between Task and Age group nor between Condition and Age group, however whole-brain analyses showed a significant three-way interaction between Task, Condition and Age group in BOLD signal in the right insula (see **Table 3** and **Figure 4**). The mean parameter estimates in this cluster were calculated and the interaction was followed up by running 2 (Condition) x 2 (Age group) mixed rmANOVAs in each Task separately. There was an interaction between Condition and Age group in the People task ($F(1, 37) = 7.20, p = 0.011, \eta_p^2 = .16$), and in the Town task ($F(1, 37) = 8.64, p = .006, \eta_p^2 = .19$).

These interactions were further explored by comparing SO and Comparison conditions within each age group. In the People task, adolescents showed greater activation in the SO than the Comparison condition ($p = .026$), while the conditions did not differ in adults ($p = .132$). In the Town task, adolescents reversely showed greater activation in the Comparison than the SO condition ($p = .018$), while again there was no difference in adults ($p = .176$).

Insert Figure 4 here

Covariate analyses

Whole-brain analyses were performed to investigate correlations between the behavioural measure of Consistency and BOLD signal during the task. No correlation between Consistency scores and individual differences in BOLD signal in the contrasts [Comparison > SO], [Comparison/Self/Other > Vowels], [People Comparison > People SO], [Town Comparison > Town SO]) was observed. Further whole-brain analyses showed that BOLD signal in the [People > Town] and [People Comparison > People SO] contrasts was not significantly related to the perspective-taking scale of the IRI.

Discussion

We performed separate behavioural and fMRI studies aiming to disentangle general and specific processes underlying relational integration of social information between late childhood and adulthood. We found behavioural evidence for general development of social and non-social relational reasoning. We discovered similar patterns of neural activity for adolescents and adults showing domain general involvement of the fronto-parietal cortex areas associated with relational integration for both social and non-social relations, and domain specific involvement of the social brain for the manipulation of social information.

Behavioural findings

Relational integration

In Study 1, the large behavioural study, we found earlier improvements in performance with age for RT and later improvements in consistency from late childhood to adulthood. Across age groups, RTs were slower in the Comparison condition, which required relational integration, than the Self and Other conditions, which required processing of a single relations. These results are in line with the robust processing speed costs observed in paradigms comparing relational integration with simpler relational processing (Crone et al., 2009; Dumontheil, Houlton, et al., 2010). In terms of relational integration performance, results show a pattern of early improvement in RT between 11-12y and 13-14y, with no further changes at older ages (**Figure 2B**), and improvements between 13-14y and 15-17y for the consistency measure, with no improvements at older ages (**Figure 2C**). Note that verbal IQ was not matched across age groups, however the differences in verbal IQ did not directly map onto the observed age effects. In particular, there was no difference in IQ between 11-12y, 13-14y and 15-17y groups, the age range where the key developmental changes were observed.

Similarly, in Study 2, the fMRI study, RTs were slower in the Comparison than in the Self and Other conditions. Relational integration was slower in adolescents than adults. There was also a trend for poorer consistency in the adolescents, which fits with the findings of the behavioural study. The less robust behavioural results in Study 2 compared with Study 1 were likely due to the smaller sample size in the imaging study.

Overall we observed that late childhood and early adolescence are associated with poorer relational integration performance than adulthood (**Figure 2B**). This pattern fits with previous findings from visuospatial relational integration tasks, which indicate poorer relational integration accuracy in 8-12 year olds than in adults (Crone et al., 2009), improvements in accuracy between 9 and 19 year olds (Rosso, Young, Femia, & Yurgelun-Todd, 2004) and poorer combined accuracy and RT in 7-9-year-olds than in 14-17-year-olds (Dumontheil, Houlton, et al., 2010, reanalysed in Dumontheil, 2014), although note that Wendelken et al. (2011) did not find age differences in 7-18-year-olds (see Dumontheil, 2014, for review). The present study is consistent with a previous investigation of the

development of the integration of semantic information using an analogical reasoning task, which demonstrated poorer performance in 6-13-year-olds than in adults (Wright, Matlen, Baym, Ferrer, & Bunge, 2008).

Social information processing

Developmental improvements in RT for relational integration did not vary as function of the type of information. However, main effects of task across age were observed. In Study 1, RTs were faster overall for Town than people, while the opposite pattern was observed in Study 2, RTs were faster overall for the People task regardless of level of relational processing, which is in line with previous findings of benefits in performance when stimuli are social rather than symbolic (den Ouden, Frith, Frith, & Blakemore, 2005; Dumontheil, Hassan, Gilbert, & Blakemore, 2010).

Both adolescents and adults showed increased speed for social information. Similarly, although consistency was overall greater in the non-social task, domain general, not social specific, developmental changes were observed. This is at odds with studies showing increased sensitivity to social stimuli during adolescence (Foulkes & Blakemore, 2016). This might be due to the low arousal and/or affective demands of the current task, and should be explored in further studies.

Furthermore, poorer overall consistency in the social task may be due to the greater complexity and variability of people's traits compared with towns' characteristics; this difference would apply to both adolescents and adults. For example, my friend Sam may be funny sometimes, but at other times he is quite grumpy, while London is always busy.

Slight discrepancies between the findings from Study 1 and Study 2 may be explained by methodological differences. Study 1 had a larger sample than Study 2, reducing the power needed to detect developmental changes in performance. In Study 1 the adjectives were the same for both the People and Town task. In Study 2, in order to maximise the mentalising requirements of the People

task, we used different adjectives for each task, which were more directly applicable to people or to towns. The fMRI task was not self-paced. These factors may have affected the behavioural results.

Our behavioural findings provide some evidence for differential performance for social information, regardless of relational level, for both adolescents and adults. Furthermore, we provide evidence for domain general development of relational integration of simple relations which does not differ as a function of the social or non-social nature of the semantic information being processed.

Neuroimaging findings

Relational integration

Relational integration was associated with greater activation in a large bilateral fronto-parietal network including the RLPFC in both the People and Town tasks. A similar pattern of activation was observed by Raposo et al. (2011) in the Compare vs Other contrast. These results further support the involvement of RLPFC and the inferior parietal cortex in relational integration (Crone et al., 2009; Wendelken et al., 2012). By adapting the task by Raposo et al. we are able to directly compare manipulation of social and non-social information and provide evidence for domain-general recruitment of the RLPFC through adolescence and adulthood across social and non-social domains.

Overall, we did not find evidence of an interaction between relational integration and social vs non-social task: both networks were recruited in parallel for relational integration and social demands.

This parallel recruitment of the two networks is similar to that observed by Meyer et al. (2012, 2015) in their social working memory task. However, these studies showed that the MPFC activation associated with social working memory was modulated by working memory load (Meyer et al., 2012, 2015), while in the present study, MPFC activation was not modulated by the number of relations participants had to consider. Similarly, although Meyer et al. (2012) found that perspective taking on the IRI scale was positively associated with social-working memory load-dependent activity within the MPFC and posterior cingulate cortex in adults, we found no association between self-reported

perspective taking on this questionnaire and behaviour or brain activity in our task. A possible source of this difference in findings is that social cognitive load in the Meyer et al. (2012) study was higher than in the present study, with the requirement to compare two, three or four individuals, whose names were maintained in working memory, on given personality traits. In the present study, our participants only compared two individuals and the relevant information remained on the screen over the duration of the trial. These differences may have contributed to both the lack of observed association between performance and MPFC activation and as well as with self-reported real-life perspective taking.

We did not observe developmental differences in activation in the RLPFC during relational integration across tasks in whole-brain analyses. Developmental differences have been previously reported in this brain region (e.g. Crone et al. 2009; Dumontheil, Houlton, et al., 2010). However, in the study by Crone et al., the age effects were mainly due to differences in time course of activations evident in their event-related design, which our block design did not allow us to test. In our previous study, age group effects were observed in ROI, vs whole brain, analyses (Dumontheil, Houlton, et al., 2010). Relational reasoning studies have traditionally involved demanding visuospatial reasoning tasks, such as the Raven Progressive Matrices. It might be that semantic reasoning about traits of people and towns is not demanding enough to tax adolescents in the same way.

Our results suggest domain-general recruitment of the RLPFC through adolescence and adulthood, independent of whether the information being manipulated is social or non-social. These results are therefore in line with the finding that the integration of visuospatial or semantic relations elicits similar activation of the relational integration network (Wendelken et al., 2012).

Social vs non-social information processing

Social information was associated with greater activation in the precuneus and MPFC in both adolescents and adults. These results are in line with a large body of literature that documents the

involvement of these regions when processing social information (e.g. see Van Overwalle, 2009, for meta-analysis). Note that, although towns can be considered social to some extent, as one can imagine the population of individuals living there, characteristics of towns were considered to be less social than traits of people.

Similar to the present study, the investigation of social relational reasoning in adults by Raposo and colleagues (2011) reported greater MPFC activation in Self, Other, and Relational conditions when contrasted to a vowel judgement condition. However, in contrast to our findings, Raposo et al. (2011) observed activation across ventral and dorsal MPFC in the Other vs Self contrast, and no activation in the Self > Other contrast. This difference between the two studies may be due to the precise question participants were asked to answer. In the present study, participants simply rated how funny (or other adjectives) they were, or their friend was, while in the Raposo et al. (2011) study participants always rated how *pleasant* or *unpleasant* they found a concept (e.g. tower), or how *pleasant* or *unpleasant* they thought their friend would find this concept. It is possible that this latter question elicited greater mentalising by asking participants to put themselves in their friend's shoes rather than asking their own opinion. Overall, our results suggest social-specific recruitment of the MPFC during adolescence and adulthood for social information, whether simple relations or integration across levels.

Interaction between Task, Condition and Age group

One neural difference between age groups was observed. At the cluster-corrected level, the right anterior insula showed a significant three way interaction between Condition, Task and Age group, driven by differential recruitment according to the domain (social vs. non-social) and the relational integration requirements with age. Although these results were not significant with an FWE-corrected threshold at the voxel level, they replicate the decreased activation with age in the anterior insula observed in a visuospatial relational reasoning task (Dumontheil, Houlton, et al., 2010). Functional changes in the anterior insula might reflect the maturation of neurocognitive

strategies, which possibly include changes in task-specific connectivity between brain regions (e.g. see Bazargani et al., 2014; Dumontheil et al., 2010; and Dumontheil, 2014 for discussion). This neuroimaging finding does not directly map onto behavioural differences between the age groups. In a previous study, we similarly observed that developmental changes in anterior insula activation during relational reasoning were not accounted for by individual differences in performance on the task (Dumontheil, Houlton, et al., 2010). Beyond differences in the sensitivity of behavioural and brain imaging measures (e.g. evidenced by greater sensitivity of neuroimaging data than behavioural data to genetic differences; Dumontheil, Roggeman, Ziermans, Peyrard-Janvid, Matsson, Kere, & Klingberg, 2010), differences may be due to the fact that behaviour reflects a large combination of factors beyond the block-related activations measured in the current fMRI paradigm, such as event-related activations, which may have had compensatory effects on performance.

Conclusion

We aimed to investigate the development of performance in social and non-social relational reasoning and their associated neural substrates. The paradigm required participants to make first and second order relational judgements about social and non-social information. Data from a behavioural and an fMRI study demonstrated development of social and non-social relational reasoning in adolescence. These behavioural results with semantic stimuli are in line with previous research using mostly visuospatial relational reasoning tasks. We did not find evidence of differential development of relational integration of social vs non-social information in our behavioural studies. Similarly, the fMRI data showed that in both adolescents and adults, relational integration of social and non-social information recruited a similar fronto-parietal network. The processing of social information additionally engaged MPFC and precuneus regions of the social brain, regardless of the order of reasoning. These findings provide further evidence that relational integration is a domain-general process (Wendelken et al., 2012).

Acknowledgements

Sarah K.G. Jensen assisted with data collection. L.M.W is funded by the Mexican National Council on Science and Technology (CONACYT). S.J.B. is funded by a Royal Society University Research Fellowship, the Wellcome Trust and the Jacobs Foundation. This research was supported by the Leverhulme Trust.

Reference list

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, *30*(3), 1059-1068. doi: 10.1016/j.neuroimage.2005.10.026
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268-277. doi: 10.1038/nrn1884
- Bazargani, N., Hillebrandt, H., Christoff, K., & Dumontheil, I. (2014). Developmental changes in effective connectivity associated with relational reasoning. *Human Brain Mapping*, *35*(7), 3262-3276. doi: 10.1002/hbm.22400
- Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267-277. doi: 10.1038/nrn2353
- Blakemore, S.-J. (2012). Development of the social brain in adolescence. *Journal of the Royal Society of Medicine*, *105*(3), 111-116. doi: 10.1258/jrsm.2011.110221
- Blakemore, S.-J., & Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nature Neuroscience*, *15*(9), 1184-1191. doi: 10.1038/nn.3177
- Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox*. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan.
- Bunge, S. A., Helskog, E. H., & Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *NeuroImage*, *46*(1), 338-342. doi: 10.1016/j.neuroimage.2009.01.064
- Bunge, S. A., Wendelken, C., Badre, D., & Wagner, A. D. (2005). Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. *Cerebral Cortex*, *15*(3), 239-249. doi: 10.1093/cercor/bhh126
- Burnett, S., & Blakemore, S.-J. (2009). The development of adolescent social cognition. *Annals of the New York Academy of Sciences*, *1167*, 51-56. doi: 10.1111/j.1749-6632.2009.04509.x
- Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development?. *Social development*, *11*(1), 1-10.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*(5), 1136-1149. doi: 10.1006/nimg.2001.0922
- Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*(9), 636-650. doi: 10.1038/nrn3313
- Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55-66. doi: 10.1111/j.1467-7687.2008.00743.x
- D'Argembeau, A., Ruby, P., Collette, F., Degueldre, C., Balteau, E., Luxen, A., . . . Salmon, E. (2007). Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of Cognitive Neuroscience*, *19*(6), 935-944. doi: 10.1162/jocn.2007.19.6.935
- David, N., Aumann, C., Santos, N. S., Bewernick, B. H., Eickhoff, S. B., Newen, A., . . . Vogeley, K. (2008). Differential involvement of the posterior temporal cortex in mentalizing but not

- perspective taking. *Social Cognitive and Affective Neuroscience*, 3(3), 279-289. doi: 10.1093/scan/nsn023
- David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., . . . Vogeley, K. (2006). Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience*, 18(6), 898-910. doi: 10.1162/jocn.2006.18.6.898
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, (10), 85.
- Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Sciences*, 7(12), 527-533. doi: 10.1016/j.tics.2003.10.004
- den Ouden, H. E., Frith, U., Frith, C., & Blakemore, S.-J. (2005). Thinking about intentions. *NeuroImage*, 28(4), 787-796. doi: 10.1016/j.neuroimage.2005.05.001
- Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostral lateral prefrontal cortex. *Developmental Cognitive Neuroscience*, 10C, 57-76. doi: 10.1016/j.dcn.2014.07.009
- Dumontheil, I., & Blakemore, S.-J. (2012). Social cognition and abstract thought in adolescence : The role of structural and functional development in rostral prefrontal cortex. *British Journal of Educational Psychology Monograph Series II*(8), 99-113
- Dumontheil, I., Burgess, P. W., & Blakemore, S.-J. (2008). Development of rostral prefrontal cortex and cognitive and behavioural disorders. *Developmental Medicine and Child Neurology*, 50(3), 168-181. doi: 10.1111/j.1469-8749.2008.02026.x
- Dumontheil, I., Hillebrandt, H., Apperly, I., & Blakemore, S.-J. (2012). Developmental differences in the control of action selection by social information. *Journal of Cognitive Neuroscience*, 24(10), 2080-2095. doi: 10.1162/jocn_a_00268
- Dumontheil, I., Houlton, R., Christoff, K., & Blakemore, S.-J. (2010). Development of relational reasoning during adolescence. *Developmental Science*, 13(6), F15-24. doi: 10.1111/j.1467-7687.2010.01014.x
- Dumontheil, I., Jensen, S. K. G., Wood, N. W., Meyer, M. L., Lieberman, M. D., & Blakemore, S. J. (2014). Preliminary investigation of the influence of dopamine regulating genes on social working memory. *Social Neuroscience*, 9(5), 437-451. doi: 10.1080/17470919.2014.925503
- Dumontheil, I., Roggeman, C., Ziermans, T., Peyrard-Janvid, M., Matsson, H., Kere, J., & Klingberg, T. (2011). Influence of the COMT genotype on working memory and brain activity changes during development. *Biological Psychiatry*, 70, 222-229. doi: 10.1016/j.biopsych.2011.02.027
- Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid Reasoning and the Developing Brain. *Frontiers in Neuroscience*, 3(1), 46-51. doi: 10.3389/neuro.01.003.2009
- 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), 459-473. doi: 10.1098/rstb.2002.1218
- Gilbert, S. J., Spengler, S., Simmons, J., Steele, J., Lawrie, S., Frith, C. D., & Burgess, P. W. (2006). Functional Specialization within Rostral Prefrontal Cortex (Area 10): A Meta-analysis. *Journal of Cognitive Neuroscience*, 18(6), 932-948. doi: 10.1162/jocn.2006.18.6.932
- Gusnard, D. A. (2005). Being a self: considerations from functional imaging. *Consciousness and Cognition*, 14(4), 679-697. doi: 10.1016/j.concog.2005.04.004
- Herting, M. M., Maxwell, E. C., Irvine, C., & Nagel, B. J. (2012). The impact of sex, puberty, and hormones on white matter microstructure in adolescents. *Cerebral Cortex*, 22(9), 1979-1992. doi: 10.1093/cercor/bhr246
- Kilford, E. J., Dumontheil, I., Wood, N. W., & Blakemore, S.-J. (2015). Influence of COMT genotype and affective distractors on the processing of self-generated thought. *Social Cognitive and Affective Neuroscience*, 10(6), 777-782. doi: 10.1093/scan/nsu118

- Kilford, E.J., Garrett, E. & Blakemore, S-J. (2016). The Development of Social Cognition in Adolescence: An Integrated Perspective. *Neuroscience & Biobehavioral Reviews*. 70, 106-120.
- Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302(5648), 1181-1185. doi: 10.1126/science.1088545
- Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cerebral Cortex*, 12(5), 477-485. doi: 10.1093/cercor/12.5.477
- Lynn, R. & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence* 32, 481-498. <http://dx.doi.org/10.1016/j.intell.2004.06.007>
- Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., & Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, 109(6), 1883-1888. doi: 10.1073/pnas.1121077109
- Meyer, M. L., Taylor, S. E., & Lieberman, M. D. (2015). Social working memory and its distinctive link to social cognitive ability: an fMRI study. *Social Cognitive and Affective Neuroscience*, 10(10), 1338-1347. doi: 10.1093/scan/nsv065
- Mills, K.L., Lalonde, F., Clasen L.S., Giedd, J.N., & Blakemore, S.J. (2014). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience*, 9(1), 123-131. doi: 10.1093/scan/nss113
- Rameson, L. T., Satpute, A. B., & Lieberman, M. D. (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage*, 50(2), 701-708. doi: 10.1016/j.neuroimage.2009.12.098
- Raposo, A., Vicens, L., Clithero, J. A., Dobbins, I. G., & Huettel, S. A. (2011). Contributions of frontopolar cortex to judgments about self, others and relations. *Social Cognitive and Affective Neuroscience*, 6(3), 260-269. doi: 10.1093/scan/nsq033
- Raven, J. C. (1998). *Manual for Raven's Progressive Matrices* (O. P. Press Ed.). Oxford.
- Raznahan, A., Shaw, P., Lalonde, F., Stockman, M., Wallace, G. L., Greenstein, D., . . . Giedd, J. N. (2011). How does your cortex grow? *Journal of Neuroscience*, 31(19), 7174-7177. doi: 10.1523/JNEUROSCI.0054-11.2011
- Rosso, I. M., Young, A. D., Femia, L. A., & Yurgelun-Todd, D. A. (2004). Cognitive and emotional components of frontal lobe functioning in childhood and adolescence. *Annals of the New York Academy of Sciences*, 1021, 355-362. doi: 10.1196/annals.1308.045
- Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience*, 4(5), 546-550. doi: 10.1038/87510
- Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of Cognitive Neuroscience*, 16(6), 988-999. doi: 10.1162/0898929041502661
- Smith, R., Keramatian, K., & Christoff, K. (2007). Localizing the rostrolateral prefrontal cortex at the individual level. *NeuroImage*, 36(4), 1387-1396. doi: 10.1016/j.neuroimage.2007.04.032
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1), 273-289. doi: 10.1006/nimg.2001.0978
- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, 30(3), 829-858. doi: 10.1002/hbm.20547
- Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience*, 16(5), 817-827. doi: 10.1162/089892904970799
- Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence (WASI). San Antonio, TX: Psychological Corporation.

- Wendelken, C., & Bunge, S. A. (2009). Transitive inference: distinct contributions of rostralateral prefrontal cortex and the hippocampus. *Journal of Cognitive Neuroscience*, *22*(5), 837-847.
- Wendelken, C., Chung, D., & Bunge, S. A. (2012). Rostrolateral prefrontal cortex: domain-general or domain-sensitive? *Human Brain Mapping*, *33*(8), 1952-1963. doi: 10.1002/hbm.21336
- Wendelken, C., Ferrer, E., Whitaker, K. J., & Bunge, S. A. (2015). Fronto-parietal network reconfiguration supports the development of reasoning ability. *Cerebral Cortex*. doi: 10.1093/cercor/bhv050
- Wendelken, C., Nakhavenko, D., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2008). "Brain is to thought as stomach is to ??": investigating the role of rostralateral prefrontal cortex in relational reasoning. *Journal of Cognitive Neuroscience*, *20*(4), 682-693. doi: 10.1162/jocn.2008.20055
- Wendelken, C., O'Hare, E. D., Whitaker, K. J., Ferrer, E., & Bunge, S. A. (2011). Increased functional selectivity over development in rostralateral prefrontal cortex. *Journal of Neuroscience*, *31*(47), 17260-17268. doi: 10.1523/jneurosci.1193-10.2011
- Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nature Reviews Neuroscience*, *4*(2), 139-147. doi: 10.1038/nrn1033
- Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience*, *1*, 8. doi: 10.3389/neuro.09.008.2007
- Zysset, S., Huber, O., Samson, A., Ferstl, E. C., & von Cramon, D. Y. (2003). Functional specialization within the anterior medial prefrontal cortex: a functional magnetic resonance imaging study with human subjects. *Neuroscience letters*, *335*(3), 183-186.

Tables

Table 1: Demographics of participants in the behavioural (Study 1) and fMRI studies (Study 2).

Age group	N	Gender	Age		Verbal IQ	
		F/M	Range	Mean (SD)	Range	Mean (SD)
Behavioural study (Study 1)						
11-12y	66	39/27	11.10-12.98	12.05 (0.51)	87-137	114.4 (12.0)
13-14y	57	30/27	13.06-14.97	13.92 (0.58)	84-140	113.1 (12.7)
15-17y	42	20/27	15.01-17.99	16.18 (0.92)	84-134	114.0 (12.4)
20-22y	43	20/22	20.34-22.97	21.93 (0.69)	97-137	118.1 (9.5)
23-25y	50	28/22	23.01-25.96	24.55 (0.92)	81-137	113.7 (13.7) ^a
26-28y	39	20/19	26.06-28.68	27.30 (0.81)	84-129	107.6 (12.2) ^b
29-39y	28	10/18	29.00-39.39	33.22 (2.80)	94-129	115.5 (9.5)
FMRI study (Study 2)						
10-16y	19	Females only	10.98-16.83	14.10 (1.89)	93-134	116.8 (11.4)
22-31y	20	Females only	22.22-31.67	25.89 (2.76)	107-131	119.9 (6.7) ^c

^a WASI data were missing for one participant.

^b WASI data were missing for three participant.

^c WASI data was missing for one adult participant.

Table 2: Mean median RT and SE (ms) for each Task (People, Town) and Condition (Self, Other, Comparison, Vowels) in Study 2.

	People		Town	
	<i>Adolescents</i>	<i>Adults</i>	<i>Adolescents</i>	<i>Adults</i>
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
Self	2215 (110)	2057 (107)	2333 (100)	2115 (98)
Other	2123 (94)	2112 (92)	2378 (96)	2171 (94)
Comparison	2570 (108)	2460 (106)	2904 (136)	2618 (133)
Vowels	2503 (112)	2212 (109)	2529 (120)	2309 (117)

Table 3: Summary of neuroimaging results. Coordinates and Z-scores are listed for regions showing a significant difference in BOLD signal for the main effect of Condition [Comparison > SO], the main effect of Task [People > Town] or [Town > People], and the interaction between Condition, Task and Age Group [(Adolescents > Adults [(Town Comparison > SO) > [People (Comparison > SO)]]). Regions labelling was done using Automatic Anatomic Labelling (AAL) (Tzourio-Mazoyer et al., 2002). BA labelling of peak of activations was done using MRICron. ^a indicates voxels where $p_{FWE} < .05$ at the voxel-level, ^b indicates clusters where $p_{FWE} < .05$ at the cluster-level, with a cluster-defining threshold of $p < .001$ uncorrected at the voxel-level. BA = Brodmann area, L/R = Left/Right hemisphere, Pre-SMA = pre-supplementary motor area.

	L/R	BA	MNI (x y z)	Z-score	Cluster size
<i>Main effect of Condition (Comparison > SO)</i>					
Lingual gyrus	L	12	-9 -85 -14	>8 ^a	4959 ^b
Calcarine gyrus		17	0 -85 1	>8 ^a	
Lingual gyrus	L	18	-21 -79 -14	>8 ^a	
Lingual gyrus	R	18	18 -79 -14	>8 ^a	
Precuneus	R	7	3 -61 46	>8 ^a	
Middle occipital gyrus	R	19	30 -79 19	7.31 ^a	
Middle occipital gyrus	L	19	-30 -76 22	6.53 ^a	
Middle temporal gyrus	L	21	-48 -46 10	6.22 ^a	
Inferior parietal gyrus	L	40	-45 -46 43	6.13 ^a	
Fusiform	R	19	27 -64 -5	5.90 ^a	
Inferior frontal gyrus	L	48	-48 14 25	6.59 ^a	1351 ^b
Precentral and middle frontal gyri	L	6	-39 -1 55	5.78 ^a	
Middle cingulate cortex and pre-SMA	R	32	9 20 46	5.77 ^a	
Inferior frontal gyrus	L	48	-36 20 22	5.68 ^a	
Middle frontal gyrus	R	44	30 14 43	5.55 ^a	819 ^b
Inferior frontal gyrus	R	44	48 26 31	5.32 ^a	
Middle frontal gyrus	R	8	30 23 52	5.03 ^a	
Middle and superior frontal gyri	R	10	30 59 7	4.56 ^a	

Precentral gyrus	R	6	30 -4 46	4.44	
Inferior and middle frontal gyri	L	47	-39 47 -2	4.91 ^a	95 ^b
Main Effect of Task (People > Town)					
Medial prefrontal cortex	R	10	6 53 13	7.22 ^a	1649 ^b
Anterior cingulate cortex	L	32	-3 53 13	7.11 ^a	
Anterior cingulate cortex	L	10	-6 44 1	5.95 ^a	
Anterior insula	R	48	30 17 -17	5.74 ^a	69
Precuneus and posterior cingulate cortex	R	23	6 -52 28	5.26 ^a	132 ^b
Main Effect of Task (Town>People)					
Lingual gyrus	R	30	9 -49 4	>8 ^a	2164 ^b
Calcarine gyrus	R	30	15 -52 13	7.65 ^a	
Calcarine gyrus	L	30	-12 -55 10	7.17 ^a	
Middle occipital gyrus	L	19	-33 -76 28	6.08 ^a	
Middle occipital gyrus	R	19	36 -70 37	5.90 ^a	
Superior occipital gyrus	L	23	-21 -64 28	5.25 ^a	
Superior occipital gyrus	R	7	24 -76 46	4.62 ^a	
Fusiform and parahippocampal gyri	L	37	-30 -37 -14	>8 ^a	192 ^b
Fusiform gyrus	R	37	30 -31 -17	6.97 ^a	302 ^b
Inferior temporal gyrus	R	20	54 -46 -11	4.13	
Parahippocampal gyrus	R	35	21 -13 -20	3.73	
Middle cingulate cortex	R	23	9 -34 34	4.78 ^a	60
Precentral gyrus and inferior frontal operculum	L	44	-42 8 28	4.58 ^a	181 ^b
Postcentral and precentral gyri	R	43	60 -10 31	3.93	108 ^b
Interaction Adolescents > Adults [(Town Comparison > SO) > [People (Comparison > SO)]					
Anterior insula	R	48	33 17 -8	4.26	178 ^b
Anterior insula	R	47	39 26 4	3.71	
Temporal pole	R	38	45 14 -20	3.59	

Figures

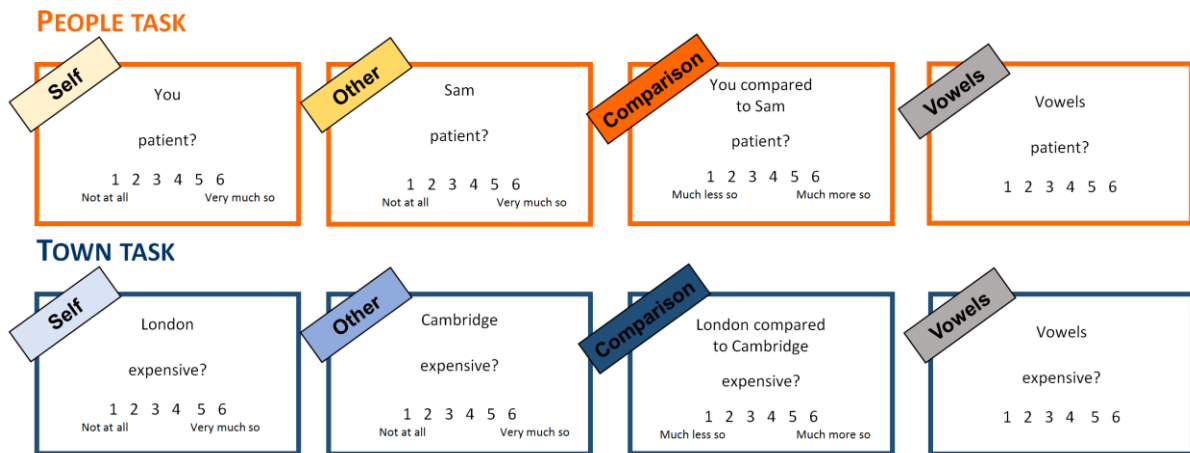


Figure 1: Example of stimuli for each task and condition. The Vowels condition was only included in Study 2. In the People task, participants were asked: “How much do you think the following words apply to you/your friend?” (Self/Other), or “How much do you think the following words apply to you compared to your friend?” (Comparison). On the rating scale, 1 indicated not at all and 6 indicated very much so in the Self and Other conditions, while 1 indicated much less so and 6 indicated much more so in the Comparison condition. All text was presented in white on a black background.

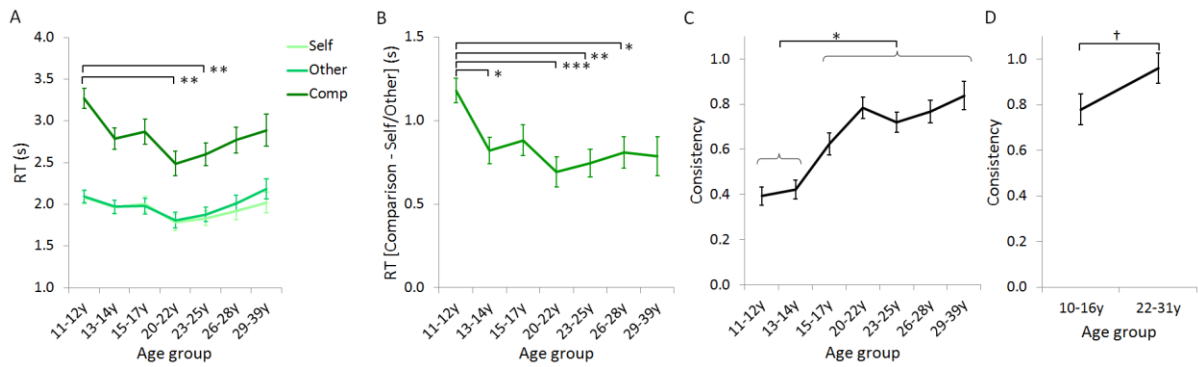
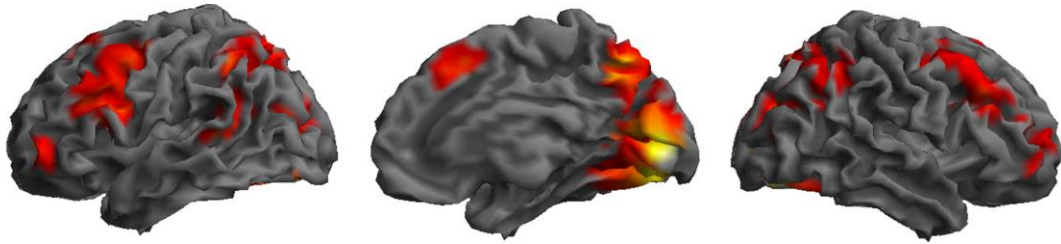


Figure 2: Mean RT and consistency scores as a function of age group. **(A)** Study 1: Mean RTs as a function of Age group and Condition. **(B)** Study 1: Mean difference in RTs between the Comparison and Self and Other conditions as a function of Age group. **(C)** Study 1: Mean consistency scores as a function of Age group. **(D)** Study 2: Mean consistency scores as a function of Age group. Error bars represent SE. * $p < .05$, ** $p < .01$, *** $p < .001$ (Bonferroni corrected).

A Comparison > Self & Other



B People > Town

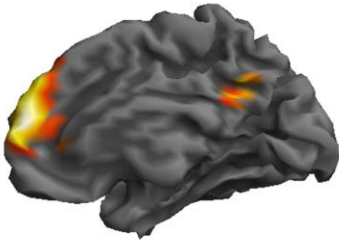


Figure 3: FMRI results across age groups. **(A)** Main effect of Condition. Regions showing increased BOLD signal in Comparison (2-REL) compared to Self and Other (1-REL) conditions are rendered on the SPM8 surface mesh template. From left to right: lateral view of the left hemisphere, medial and lateral views of the right hemisphere. **(B)** Main effect of Task. Regions showing increased BOLD signal in the People compared to the Town task are rendered on the SPM8 surface mesh template (medial view of the right hemisphere).

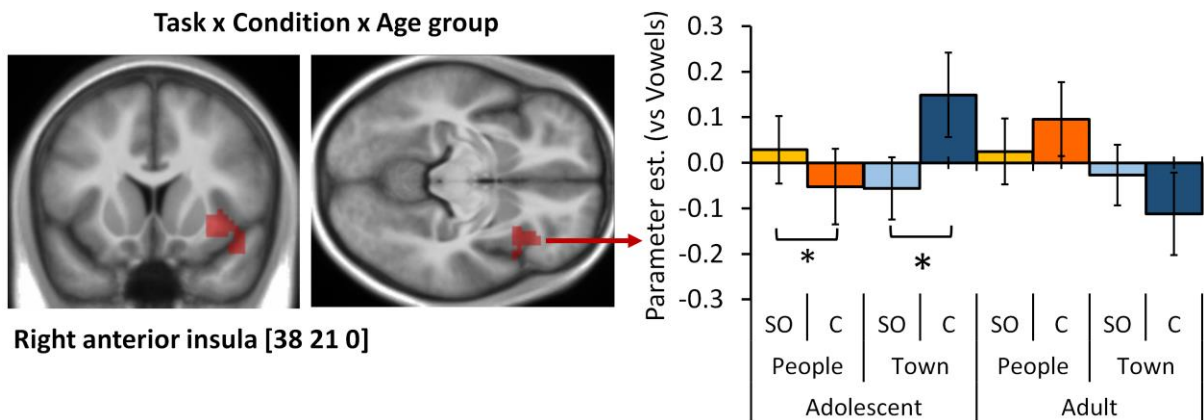


Figure 4: FMRI results of the interaction between Task, Condition and Age group. On the left, the right anterior insula cluster showing a three-way interaction between Task, Condition and Age group is shown on an average structural scan of all participants in the study (left $y = 21$; right: $z = 0$). The contrast was thresholded at $p < .001$ uncorrected at the voxel level, $p_{FWE} < .05$ at the cluster level. On the right, parameter estimates extracted from this cluster are plotted in a bar chart to illustrate the three-way interaction. Error bars represent SE. † $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix

List of stimuli for each task in Study 2. In the People task, participants were asked: “How much do you think the following words apply to you/your friend?”, or “How much do you think the following words apply to you compared to your friend?”. In the Town task, participants were asked: “How much do you think the following words apply to your town/other town?”, or “How much do you think the following words apply to your town compared to the other town?”. Familiarity and frequency measures were included to ensure that all adjectives were commonly used English words, and that their occurrence was comparable between Tasks.

	People	Town	
Adjectives	arrogant, jealous, timid, selfish, careless, witty, cheerful, thoughtful, stubborn, ambitious, confident, aggressive, clever, smart, bold, brave, generous, helpful, mature, wise, tough, funny, curious, honest, sensitive, friendly, fair, patient, bright, happy	run-down, shabby, boring, rainy, noisy, sleepy, vibrant, picturesque, polluted, quaint, historic, lively, dull, exciting, romantic, dirty, urban, expensive, dangerous, rural, cultural, safe, unusual, clean, quiet, traditional, famous, amazing, flat, ancient	
Number of letters	$M = 6.7, SD = 1.8,$	$M = 6.6, SD = 1.9$	$t(58) = 0.14, p = .89$
Number of vowels	$M = 2.5, SD = 1.1$	$M = 2.6, SD = 1.1$	$t(58) = 0.36, p = .73$
Familiarity	$M = 541.6, SD = 48.6$	$M = 561.6, SD = 53.6$	$t(40) = 1.2, p = .23$
Brown frequency	$M = 9.26, SD = 10.6$	$M = 7.70, SD = 11.0$	$t(41) = 0.47, p = .64$
Kucera-Francis frequency	$M = 32.0, SD = 26.9$	$M = 35.3, SD = 27.2$	$t(58) = 0.47, p = .64$