doi:10.1093/scan/nsu030 SCAN (2015) 10, 114–121

Neural correlates of attributing causes to the self, another person and the situation

Jenny Kestemont, Ning Ma, Kris Baetens, Nikki Clément, Frank Van Overwalle, and Marie Vandekerckhove
Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

This study compares brain activation during causal attribution to three different loci, the self, another person and the situation; and further explores correlations with clinical scales (i.e. depression, anxiety and autism) in a typical population. While they underwent functional magnetic resonance imaging, 20 participants read short sentences about another person ('someone') who engaged in behaviors with the participant or made comments about the participant. The participants then attributed these behaviors to three attribution loci: themselves, the other person or the situation. The results revealed common activation across the three attribution loci in the bilateral temporo-parietal junction (TPJ), left posterior superior temporal sulcus, precuneus and right temporal pole (TP). Comparisons between the attribution loci revealed very little differences, except for increased activation of the right TP while making attributions to the situation compared with the self. In addition, when making attributions to the situation or other persons for negative events, there were reliable correlations between low activity in the left TPJ and high levels of anxiety and problematic social interaction in autism. The results indicate that attributions to different loci are based on the same underlying brain process, which might be atypical among persons with anxiety or autism symptoms.

Keywords: causal attribution; fMRI; psychopathology; autism; anxiety

INTRODUCTION

Identifying causes of human behavior within the self, another person or situational circumstances enables us to understand what is going on in our social interaction, and provides a guideline for future contact. To categorize these causes requires us to understand not only the situational context but also the psychology of other people's mind: their intention, aims and thoughts. This knowledge on others' mental states is called *mentalizing*. Attributing causality is, therefore, highly dependent on mentalizing capacities that promote adequate social understanding and interaction of healthy individuals. Impairment of these mentalizing capacities leads to faulty causal attribution tendencies and is inherent to clinical conditions such as depression, anxiety and autism (Arkin *et al.*, 1980; Hope *et al.*, 1989; Dykema *et al.*, 1996; Kinderman and Bentall, 1996, 1997; Craig *et al.*, 2004). The goal of this research is to explore brain areas that support causal attribution, and to analyze which of these areas are associated with mentalizing and correlate with clinical scales measuring depression, anxiety and autism in a healthy population.

Causal attribution and the brain

According to Kelley (1973), causes of observed behavior are typically attributed to the agent who acts out the behavior, the observer's own behavior or feelings or to situational circumstances of the context. Distinguishing between causes that are internal (i.e. self) or external (i.e. other people or circumstances) has been termed the *locus* dimension of causality (Russel, 1982; Weiner, 1985). However, a limitation of current neuroimaging research on causal attribution is that comparisons have been made only between self and other (Farrer and Frith, 2002), between self and external causes (collapsing across another person and the situation; Blackwood *et al.*, 2003; Seidel *et al.*, 2010), between other and situational causes (ignoring the self; Kestemont *et al.*, 2013) and one study only examined other person attributions (Harris *et al.*, 2005). To avoid this limitation, in this study, we classify causes along the three possible loci of the locus dimension, that is, the self, the other person or the situation, and investigate the neural correlates of each attribution locus separately and in comparison with each other. We make use of a recent self-report questionnaire, the *Internal, Personal and Situational Attributions Questionnaire* (IPSAQ) designed and developed by Kinderman and Bentall (1996), which allows causes to be classified along the three loci of the locus dimension, and so improves on the earlier 2-fold self-external distinctions used in previous studies.

According to recent meta-analyses (Van Overwalle, 2009; Mar, 2011; Denny *et al.*, 2012; Schilbach *et al.*, 2012), mentalizing or understanding the causes of social behavior, recruits a network of midline and temporal brain areas. According to Van Overwalle (2009), some of these areas are responsible for the understanding of temporary or here-and-now behaviors and beliefs in the current situation, including the temporo-parietal junction (TPJ), and the precuneus (PC), while the posterior superior temporal sulcus (pSTS) provides amodal sensory input to these mentalizing areas. As this study is set in the context of single, temporary events, these areas are especially relevant in this study. In addition, the medial prefrontal cortex (mPFC) is responsible for the identification of stable and abstract characteristics of persons such as traits (see also Harris *et al.*, 2005; Lieberman, 2007; Carrington and Bailey, 2009; Mitchell, 2009; Ma *et al.*, 2012). In addition, the temporal poles (TPs) are sometimes recruited during mentalizing, and are believed to be involved in memory-related processing of social information (Olson *et al.*, 2007; Ross and Olson, 2010).

In a recent functional magnetic resonance imaging (fMRI) study on causal attributions (Kestemont *et al.*, 2013), participants read short descriptions of behaviors or events and made attributions to the agent (i.e. another person) or the situation. Scanning revealed activation in brain areas typically involved in mentalizing about single, temporary events (e.g. Kornap gets a present), including the TPJ, pSTS and PC. Interestingly, stronger activation of these mentalizing areas was found in attributions to the situation compared with the person (Kestemont *et al.*, 2013). A number of studies reported a similar increase of activation for external attributions in the TPJ and PC (Farrer and Frith, 2002; Seidel *et al.*, 2010) and pSTS (Blackwood *et al.*, 2003),

Received 5 July 2013; Revised 17 January 2014; Accepted 10 February 2014
Advance Access publication 13 March 2014

This research was funded by GOA68 and SRP15 grants awarded by the Vrije Universiteit Brussels, Belgium. This research was supported by an OZR Grant of the Vrije Universiteit Brussel to F.V.O. and performed at Ghent Institute for Functional and Metabolic Imaging (GIfMI).

Correspondence should be addressed to Jenny Kestemont, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel, Belgium. E-mail: Jenny.Kestemont@vub.ac.be

although they did not make a fine-grained distinction between external persons and situations. These results are consistent with the notion that attributions to the other person are often made spontaneously while situation attributions require more elaborate processing, in line with behavioral research (cf. *fundamental attribution bias*; Gilbert *et al.*, 1988).

Clinical symptoms and attributional biases

One of the major advantages of a self-report methodology as we use here is that a person's subjective interpretation of the causal locus of an event is an important predictor of psychological well-being and future behavior (Peterson *et al.*, 1982; Russel, 1982; Weiner, 1985; Van Overwalle, 1989; Dykema *et al.*, 1996). Personal styles or biases in attributional thinking may be indicative of maladaptive functioning and psychopathology (Buchanan and Seligman, 1995).

A plethora of behavioral studies has documented that healthy people tend to protect their self-esteem and psychological well-being by showing a *self-serving bias*, or the tendency to make attributions to the self after positive events, while blaming external factors for negative events (Miller and Ross, 1975; Peterson *et al.*, 1982; Van Overwalle, 1989; Kinderman and Bentall, 1996). In contrast, clinical groups often show a reversed bias, called a *self-blaming bias*, in which negative events are attributed to themselves and positive events externally. This self-blaming pattern is typical among depressed (Peterson *et al.*, 1982, 1985; Peterson and Seligman, 1984; Kinderman and Bentall, 1997; Diez-Alegria *et al.*, 2006; Northoff, 2007) and anxious individuals (Arkin *et al.*, 1980; Hope *et al.*, 1989), but seems to be absent among autistic individuals (Blackshaw *et al.*, 2001; Craig *et al.*, 2004). Based on these findings, it seems plausible that the neurological underpinnings of these biases might serve as diagnostic signals of these pathologies.

Neuroimaging research on attributional biases started only recently, and seems to indicate that self-blaming attributions in clinical populations—although made quite often—recruit more activity in mentalizing areas than healthy controls. For instance, increased activation in the mPFC reflecting self-blaming attributions was found for depressed (Yoshimura *et al.*, 2010, 2013; but see Seidel *et al.*, 2012) and anxious individuals (Paulesu *et al.*, 2010). However, we know of no reports of differential mentalizing activity reflecting biased attributional processing among autistic individuals.

Present research and hypotheses

The first goal of this study is to measure brain activity during attributions of a single, temporary event. We use events as described in the IPSAQ, as this material is open to personal interpretations of the implied cause, which makes it more sensitive to deviations indicative of clinical dysfunctioning. Contrary to prior research, we distinguish not only between internal *vs* external attributions but also divide attributions further up into three distinct, more fine-grained causal loci, including the self, another person and the situation. We expect that these distinct attribution loci will induce mentalizing about momentary acts or thoughts of someone else based on goals, wishes and mood states of that person at that moment. This generally recruits activation in areas involved during mentalizing about temporary or here-and-now events (cf. pSTS, TPJ, PC; Van Overwalle, 2009) rather than the mPFC because activity in this area typically reflects stable (person) attributions such as traits (Van Overwalle, 2009). Moreover, based on our earlier findings (Kestemont *et al.*, 2013), we predict that activation in these mentalizing areas will be further increased during causal attributions that emphasize the situation, because this requires increased effortful processing relative to attributions made to the other person (Kestemont *et al.*, 2013).

The second goal of this study is to correlate brain activity with increasing levels of subclinical psychopathology. We expect that subclinical levels of psychopathology reveal a typical pattern of self-blaming, especially given symptoms of depression and anxiety, both at the behavioral and neurological level (i.e. increasing mentalizing activity). In contrast, our sample as a whole is expected to show a healthy, self-serving bias.

METHODS

Participants

Twenty right-handed Dutch-speaking participants were recruited for this study. Their age ranged from 18 to 41 years, with a mean age of 23.2 years. Nine of the participants were men, 11 were women. Participants were paid €10 and received a CD with their structural scanning images in exchange for their participation. The participants were recruited via university mailing lists. All the participants had normal or corrected-to-normal vision, and none of them reported any abnormal neurological history. All participants complied with the following selection criteria: no internal metal objects or artificial implants, no dental brackets or other important dentures, no increased risk for epileptic attacks, no psychiatric diagnosis and no pregnant women or women giving breastfeeding. Informed consent was obtained in a manner approved by the Medical Ethics Committee of the University Hospital Ghent (where the study was conducted) and the Vrije Universiteit Brussel (of the principal investigator).

Stimulus material

The stimulus material consisted of 80 experimental (see 'Appendix' section) and 40 baseline sentences. The experimental sentences described behaviors and thoughts of someone else involving you (e.g. Someone lies to you and Someone thinks you are smart). These sentences were the same as used in the study of Seidel *et al.* (2010) translated from German into Dutch, but the actor in Seidel's sentences ('a friend') was changed into an unspecified 'someone' to allow for a more unbiased assessment of attributions (e.g. as a self-bias might include also people close to us, such as friends). Of these sentences, 40 had a positive valence and 40 had a negative valence. Note that of the 80 experimental sentences borrowed from Seidel *et al.* (2010), 32 were based on the sentences of the IPSAQ (Kinderman and Bentall, 1996). The 40 baseline sentences described semantic statements of which participants had to judge whether they were true, false or unknown (e.g. Tokio is the main city of Japan). These sentences involved non-mental facts, which are typically used as baseline in mentalizing studies (Van Overwalle, 2009).

A pilot-study ($N = 124$) ensured that the three loci ('self', 'other person' and 'situation') were used about equally often: 28% were attributed to the self, 42% to the other person and 30% to the situation. During the fMRI experiment, the attributions were respectively 30%, 41% and 29%.

Procedure

The participants were instructed to read each event very carefully and to imagine that it happened to them, to think about a cause of the event and to categorize it as something in the self, in the other person or in the situation. Several examples and descriptions were given in pre-training sessions, to become familiar with the different tasks.

During functional scanning, each experimental trial started with the instruction (2 s): 'search cause', followed by a fixation cross (jittered between 3 and 5 s). Next, the experimental sentence was presented (5.5 s), followed by a fixation cross (0.5 s). Finally, a question appeared: 'the cause lies in:' and participants had to choose between 'self', 'other person' or 'situation' by pressing the appropriate response button.

This means that participants could assign the same experimental sentence to different causal categories. The question was presented for 7 s or until a response was given. The procedure was identical for baseline trials, except that the question was: 'Is this true?' and participants responded by pressing the appropriate button ranging from 1 = no, 2 = I don't know, to 3 = yes.

Clinical scales

After scanning, participants filled in a booklet with several questionnaires.

Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS) is a self-report questionnaire, screening for the presence of anxiety and depressive states. It consists of 14 items divided into two subscales: depression and anxiety. A validated Dutch version by Spinhoven *et al.* (1997) was used (Zigmond and Snaith, 1983).

Autism-spectrum Quotient

The questionnaire consisting of 50 items divided over two subscales (attention to detail and social interaction) is a self-report questionnaire screening the degree of development of autistic characteristics in adults. Hoekstra *et al.* (2008) translated and validated the questionnaire for a Dutch population (Baron-Cohen *et al.*, 2001).

Imaging procedure

Images were collected with a 3 T magnetrom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany), using an eight-channel radiofrequency head coil. Stimuli were projected onto a screen at the end of the magnet bore that participants viewed by a mirror mounted on the head coil. Stimulus presentation was controlled by E-Prime 2.0 (www.pstnet.com/eprime; Psychology Software Tools) under Windows XP. Foam cushions were placed within the head coil to minimize head movements. We first collected a high-resolution T1-weighted structural scan (MP RAGE) followed by one functional run of 922 volume acquisitions (30 axial slices; 4 mm thick; 1 mm skip). Functional scanning used a gradient-echo echo planer pulse sequence (repetition time = 2 s; echo-time = 33 mm; 3.5 mm × 3.5 mm × 4.0 mm resolution).

Image processing

The fMRI data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK). For each functional run, data were preprocessed to remove sources of noise and artifact. Functional data were corrected for differences in acquisition time between slices for each whole-brain volume, realigned within and across runs to correct for head movement. The functional data were then transformed into a standard anatomical space (2 mm isotropic voxels) based on the ICBM152 brain template [Montreal Neurological Institute (MNI)], which approximates Talairach and Tournoux atlas space. Normalized data were then spatially smoothed (6 mm full-width at half-maximum) using a Gaussian Kernel. Finally, realigned data were examined, using the Artifact Detection Tool software package (ART; http://www.nitrc.org/projects/artifact_detect), for excessive motion artifacts and for correlations between motion and experimental design, and between global mean signal and experimental design. Outliers were identified in the temporal differences series by assessing between-scan differences using the default criteria of ART (*Z*-threshold: 3.0 mm; scan to scan movement threshold: 0.5 mm; rotation threshold: 0.02 radians). These outliers were omitted from the analysis by including a single regressor for each outlier. No correlations between motion and experimental design or global signal and experimental design were identified. Six directions of motion parameters from the realignment step as well as outlier time points (defined by ART) were included as nuisance regressors. We used a default high-pass filter of 128 s and serial correlations were accounted for by the default auto-regressive AR(1) model.

Statistical analysis

The statistical analysis of the fMRI data involved first-level single participant analyses with a regressor for each condition time-locked at the presentation of the sentence, six movement artifact regressors, and a variable amount of artifact regressors determined by ART, and applying a canonical response function with event duration set to 0, using the general linear model of SPM8 (Wellcome Department of Cognitive Neurology). Analyses of interest were performed at the group second-level on the parameter estimates (regressors) associated with each condition using a random-effects model. The statistical analysis involved two within-participants factors: attribution locus (self, other person or situation) and valence (positive or negative event), with the truth statements as baseline. A whole-brain analysis of variance (ANOVA) failed to reveal significant main effects of attribution locus or valence, or their interaction [this was largely confirmed by a % signal change (% SC) analysis described below]. Hence, all further neuroimaging analyses were conducted omitting the valence factor.

To test our specific hypotheses, we computed contrasts between each attribution locus condition and the truth baseline, as well as their conjunction. We also tested the attributional biases using the following contrasts for the self-serving bias (positive self + negative other + negative situation) − (negative self + positive other + positive situation) and the reverse contrast for the self-blaming bias. All contrasts were first computed on a priori regions of interest (ROIs) with the small volume correction in SPM8 (Wellcome Department of Cognitive Neurology). The ROIs involved a sphere of 8 mm radius around the centers (in MNI coordinates) of areas that were identified in the meta-analysis of Van Overwalle (2009) and Van Overwalle and Baetens (2009) as involved in mentalizing: 0, −60, 40 (PC), ±50, −55, 10 (pSTS), ±50, −55, 25 (TPJ), 0, 50, 20 (mPFC), and by Sugiura *et al.* (2006) as involved in person identity, ±45, 5, −30 (TP). Next, we conducted a whole-brain analysis to identify other significant regions, using a voxel-based statistical threshold of $P \leq 0.001$ (uncorrected). For both ROI (small-volume) and whole-brain analyses, we list only those areas that survive a threshold of $P \leq 0.05$, family-wise error (FWE) corrected, and a minimum volume of 10 voxels (or 5 if there are 10 voxels in another contrast, see Table 1).

In addition, the mean % SC in each ROI was extracted using the MarsBar toolbox (http://marsbar.Sourceforge.net) for all attribution locus by valence conditions. For behavioral and imagining data (% SC), we computed correlations with the clinical questionnaires [subscales of the HADS and Autism-spectrum Quotient (AQ)]. To avoid false positives given the large number of correlations involved, we used stricter Bonferroni-corrected thresholds. First, for the behavioral data, the hypothesized correlations between clinical scales and attribution ratings were thresholded at an uncorrected level ($P < 0.05$), although further exploratory correlations were corrected at a stricter threshold of $P < 0.01$ (corrected for 3 attribution loci and 2 levels of valence; or $P < 0.05/6 \approx 0.01$). Second, for the imaging data, the correlations with clinical scales were restricted to the three most important mentalizing areas that were significant in the main analysis (bilateral TPJ and PC). The hypothesized correlations were thresholded at $P < 0.01$ (corrected for 3 attribution loci and 2 levels of valence; or $P < 0.05/6 \approx 0.01$), and further exploratory analyses were thresholded at a stricter $P < 0.005$

Table 1 Contrasts of self, other and situation (>truth baseline) and their conjunction for ROI and other regions (whole-brain analysis)

		Self > truth					Other > truth					Situation > truth					Conjunction				
Anatomical label	Brodmann	*x*	*y*	*z*	Voxels	Max *t*	*x*	*y*	*z*	Voxels	Max *t*	*x*	*y*	*z*	Voxels	Max *t*	*x*	*y*	*z*	Voxels	Max *t*
ROI																					
mPFC	9						4	56	22	15	3.87**										
R TP	21	48	8	−36	7	3.72*	48	8	−36	51	5.13***	50	4	−34	61	5.44***	48	8	−36	7	3.72*
R TPJ	13	44	−50	26	225	6.17***	44	−50	26	247	6.36***	44	−50	26	249	6.74***	44	−50	26	225	6.17***
L TPJ	39	−44	−56	30	238	6.33***	−44	−58	28	242	6.52***	−44	−56	30	230	6.26***	−44	−56	30	230	6.26***
R pSTS	22/39						46	−54	16	39	4.16**	46	−52	16	57	4.15**					
L pSTS	22	−56	−56	14	18	4.36**	−56	−56	14	48	4.91***	−56	−56	14	47	4.89***	−56	−56	14	18	4.36**
PC	7	−6	−64	40	257	7.61***	2	−60	40	257	7.87***	2	−60	40	257	7.26***	2	−60	40	257	7.26***
Other regions																					
L frontal-sup-medial (dmPFC)	9						−10	52	40	515	5.30*										
R mid-temporal	21						54	−6	−22	434	6.34**	52	−6	−24	368	6.51***					
Cingulate gyrus	23						−4	−20	28	563	6.59***	−4	−18	28	528	6.80***	−4	−20	28	341	6.59***
Posterior cingulate cortex	31	−10	−48	34	2478	8.03***	−10	−48	34	1964	8.18***	−10	−48	34	1960	7.48***	−10	−48	34	1764	7.48***
R sup-temporal	40	66	−48	24	2048	6.45***	66	−48	22	3007[d]	6.58***	66	−48	22	2520	6.75***	66	−48	22	1564	6.45***
R supramarginal	40						66	−30	28	3007[d]	6.43***										
L angular	39	−38	−56	28	2447[c]	8.48***	−38	−56	28	3299[e]	9.32***	−38	−56	28	2121[g]	9.32***	−38	−56	28	1850[h]	8.48***
L supramarginal	40	−60	−48	30	2447[c]	6.45***	−60	−48	30	3299[e]	7.05***	−62	−48	30	2121[g]	6.65***	−60	−48	30	1850[h]	6.45***
L mid-temporal	21/22	−60	−56	16	2447[c]	5.46*						−60	−58	16	2121[g]	6.01**	−60	−56	16	1850[h]	5.64*
L sub-gyral (temporal lobe)	21/22						−46	−36	−6	3299	5.95**	−46	−36	−6	484	5.41*					

Coordinates refer to the MNI stereotaxic space. ROIs are spheres with 8 mm radius around coordinates 0, −60, 40 (PC), ±50, −55, 10 (pSTS), ±50, −55, 25 (TPJ), 0, 50, 20 (mPFC) and ±45, 5, −30 (TP). R, right; L, left; dmPFC, dorsomedial prefrontal cortex. All clusters thresholded at $P < 0.001$ (for ROIs corrected after small volume analysis; for other regions corrected after whole-brain analysis) with a minimum cluster threshold of 10 (or 5 if >10 elsewhere on the same row). Only peaks that are significant ($P < 0.05$, FWE corrected) in at least one contrast are listed. Clusters with the same superscript activate more than one region.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, FWE corrected.

(corrected for 3 attribution loci, 2 levels of valence and 3 brain areas; or $P < 0.05/18 \approx 0.005$).

RESULTS

Behavioral results

To test for a self-serving bias, we compared internal (self) *vs* external attributions (other person or situation) taking valence into account. A repeated measures ANOVA revealed a main effect of attribution locus [$F(1, 19) = 117.09$, $P < 0.001$] indicating that more attributions were made to external causes than to internal causes. More importantly, in line with the self-serving bias, there was a significant interaction between attribution locus and valence [$F(1, 19) = 14.71$, $P = 0.001$] showing that this tendency was larger for negative events ($M_{external} = 30.45$ and $M_{internal} = 9.40$) than for positive events ($M_{external} = 25.30$ and $M_{internal} = 14.40$).

Correlations between clinical scales and attributions confirmed our hypothesis that participants with increased depressive symptoms (on the HADS) showed increased self-blaming in that they made more attributions of positive events to the other person ($r = 0.49$, $P < 0.05$), although they did not make more attributions of negative events to the self ($r = 0.13$, ns). Contrary to our predictions, participants with anxiety symptoms did not show the hypothesized self-blaming bias. An exploratory analysis using a stricter level of $P < 0.01$ (see 'Methods' section) further revealed that autistic symptoms of attention to detail were associated with a self-blaming trend, in that there was a negative correlation between autistic symptoms and (self-serving) attributions of negative events to another person ($r = -0.64$, $P < 0.005$).

Imaging results

Attributions to the self, another person and the situation

We predicted activity in mentalizing areas dealing with judgments of temporary events (cf. PC, pSTS and TPJ) across all attribution loci, although we also expected that activity would further increase for attributions to the situation.

As predicted, each attribution locus > truth baseline contrast showed almost identical activation patterns for attributions to the self, another person and situation. A conjunction analysis of these three comparisons confirmed this similarity across all attribution loci (Table 1; Figure 1). Consistent with our hypothesis, the conjunction revealed activation in a priori defined ROIs involving the PC, right TP, bilateral TPJ and left pSTS ($P < 0.05$, small-volume FWE corrected). Additional activations in the conjunction were revealed by the whole-brain analysis, and revealed activations often extending from the ROIs: the (posterior) cingulate cortex (adjacent to the PC), the right superior temporal gyrus (adjacent to the STS), the left angular gyrus and left supramarginal gyrus (adjacent to the pSTS and TPJ) and left medial temporal gyrus ($P < 0.05$, whole-brain FWE corrected).

We also directly compared attribution loci among each other, and found only one significant difference indicating that situation attributions generated stronger activation in the right TP than self attributions ($P < 0.05$, small-volume FWE corrected), which only partly supports our hypothesis of greater mentalizing activation during situation attributions.

Clinical scales and attributional biases

An initial analysis on our whole participant sample (see 'Methods' section) revealed no mentalizing activity reflecting a self-serving or self-blaming bias. Next, we tested our hypothesis that psychopathological symptoms are associated with increased mentalizing activity revealing self-blaming attributions, using a corrected threshold of $P < 0.01$ (see 'Methods' section). However, none of the predicted correlations was significant. We then further explored other correlations between subclinical levels of psychopathology and mentalizing activity for the most important mentalizing ROIs which were significant in the conjunction (i.e. bilateral TPJ and PC), using a strict threshold of $P < 0.005$ (see 'Methods' section). Unexpectedly, we found that activation in the left TPJ was associated with less rather than more self-blaming attributions. In particular, when making attributions of negative events to the situation or to other persons, activity in the TPJ

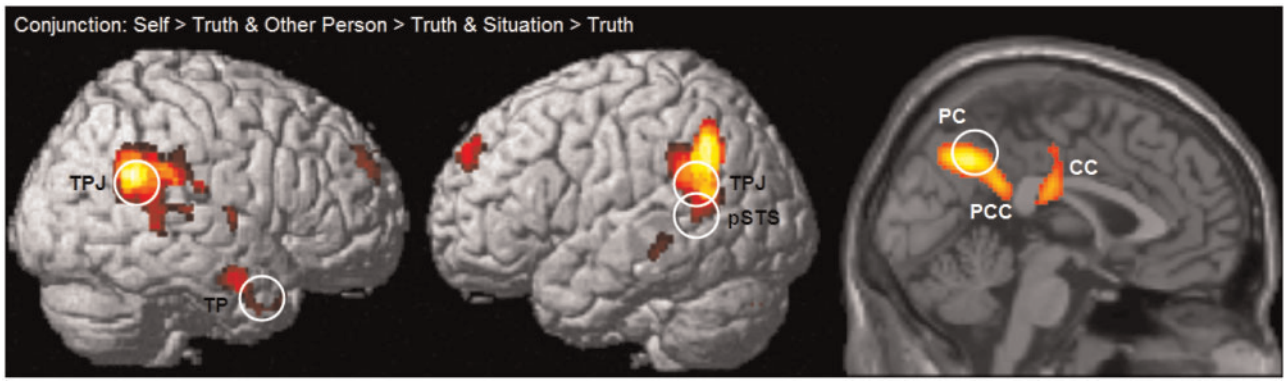

Fig. 1 Conjunction of self > truth and other person > truth and situation > truth contrasts thresholded at $P < 0.001$ (whole-brain uncorrected). Circles denote a priori ROIs.

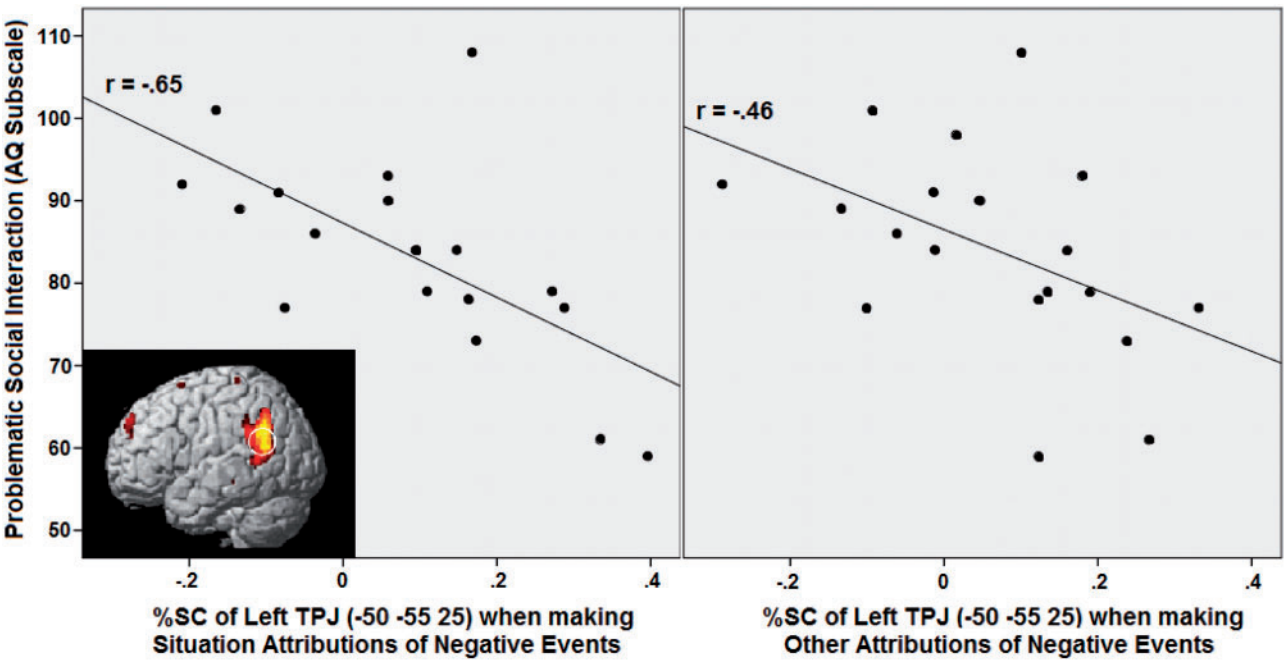

Fig. 2 Pearson correlation between the Social Interaction subscale of the AQ and %SC in the ROI of the left TPJ during presentation of negative events attributed to the situation ($P < 0.005$; left) and the other person ($P < 0.05$; right).

correlated negatively with increased scores on the Problematic Social Interaction subscale of the AQ (Figure 2; although the correlation with other person attributions, $r = -0.46$, $P < 0.05$, did not meet the strict $P < 0.005$ threshold), and on the Anxiety subscale of the HADS ($P < 0.005$; Figure 3). No other correlations surpassed the strict threshold.

DISCUSSION

This study explored the mentalizing brain areas involved in causal attributions distinguishing for the first time between three causal loci: the self, the other person or the situation. Moreover, we explored the relationship between brain activity during (biased) causal attribution and psychopathology scores on clinical scales in a typical population.

At the behavioral level, our results for our whole (healthy) sample revealed a self-serving bias, in line with the findings in the literature (e.g. Kinderman and Bentall, 1996). Moreover, we found limited support for the predicted reversal into a self-blaming bias given clinical symptoms. In particular, self-blaming was associated with depressive symptoms as reported in earlier studies (Peterson *et al.*, 1982, 1985; Peterson and Seligman, 1984; Kinderman and Bentall, 1997; Yoshimura *et al.*, 2010, 2013; Seidel *et al.*, 2012), but not with anxious symptoms, contrary to earlier studies (Arkin *et al.*, 1980; Hope *et al.*, 1989). Moreover, autistic symptoms (i.e. attention to detail) were associated with decreased self-serving attributions.

At the neural level, our imaging findings confirmed that all three attribution loci (self, another person and situation) recruited mentalizing ROIs dealing with social understanding of temporary, here-and-now, events (i.e. TPJ, pSTS and PC) as well as the TP. A whole-brain analysis revealed additional areas that were most often extensions from these ROIs. This confirms and extends research from our laboratory documenting activation of the same mentalizing areas when the self was not involved (Kestemont *et al.*, 2013). Moreover, it moves away from the limited focus of earlier research on the comparison between internal *vs* external attributions (collapsed across another person and situation), in which the separate contribution of each attribution loci was neglected (Farrer and Frith, 2002; Blackwood *et al.*, 2003; Seidel *et al.*, 2010). Consistent with predictions, as our participants were

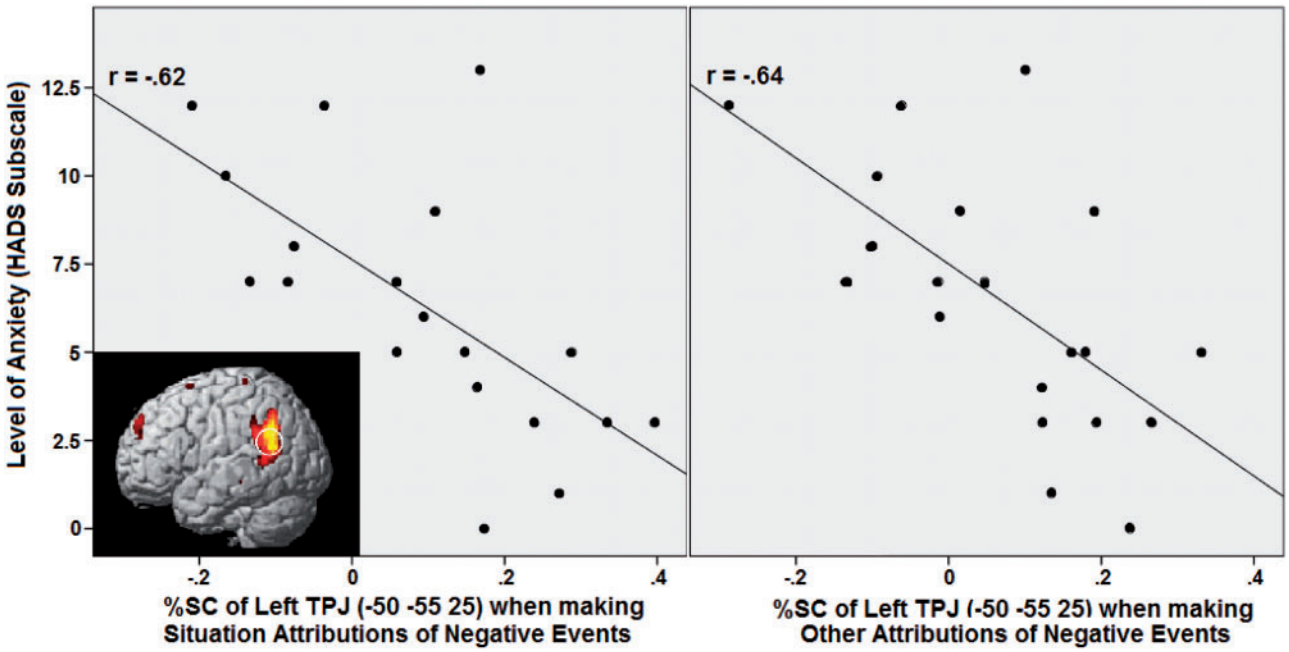

Fig. 3 Pearson correlation between the Anxiety subscale of the HADS and %SC in the ROI of the left TPJ during presentation of negative events attributed to the situation ($P < 0.005$; left) and the other person ($P < 0.005$; right).

requested to provide a cause for a temporary event rather than an enduring trait about an agent, the mPFC was not engaged (see also Ma *et al.*, 2012).

Moreover, our study shows that most differences between attribution loci reported by (Kestemont *et al.*, 2013) disappear when all loci are taken into consideration and included in a response. This might indicate that the underlying process of identifying and selecting a cause is independent of the locus (i.e. the self, another person or situation) where the cause is finally assigned to. Perhaps, briefly considering various internal or external loci is a natural process, although we cannot exclude the possibility that this overlap has been induced somewhat artificially by the experimental instruction to decide between these distinct loci. Nevertheless, we found that situation attributions generated stronger activation only in the right TP compared to self attributions, and thus might require deeper context-related memory-related processing (Olson *et al.*, 2007; Ross and Olson, 2010). In the previous study by Kestemont *et al.* (2013), situation attributions recruited mentalizing areas more broadly, as increased activity was not only found in the TP but also in the bilateral pSTS and TPJ. Perhaps, this decreased differential activation for situation attributions is due to the additional involvement of the self in this study. Behavioral research (Jones and Nisbett, 1972; Taylor and Fiske, 1975; Ross, 1977; Gilbert and Malone, 1995) has demonstrated that when the self is an active agent, the situation becomes more salient. This phenomenon is known as the actor–observer difference, and is often explained by an increased visual saliency of the environment by an active agent, who behaves in accordance to the limits set by the external situation (in contrast to a mere observer; Jones and Nisbett, 1972; Taylor and Fiske, 1975; Ross, 1977; Gilbert and Malone, 1995). Thus, by including the self, the situation might have become more prominent than in the previous study by Kestemont *et al.* (2013), reducing the differences in saliency and processing load for all three attribution loci in this study.

In line with the limited behavioral evidence for attributional biases given elevated symptoms of psychopathology, we found little neural correlates of the predicted self-blaming bias. This failure is most likely due to our selection of a non-clinical sample, where these biases are presumably less extreme, in contrast to earlier research that used clinical subsamples and biases were neurologically detectable (Blackwood *et al.*, 2003; Paulesu *et al.*, 2010; Yoshimura *et al.*, 2010, 2013; Seidel *et al.*, 2012).

Nevertheless, exploratory correlations at a corrected threshold between brain activation and clinical symptoms revealed that there were robust negative correlations between autism and anxiety scores and left TPJ activation when making self-serving attributions to the situation and other persons for negative events. Although unexpected, this pattern was systematic among anxious and autistic symptoms, for external attributions to both the person and the situation. This pattern is, therefore, of clinical importance, as it may suggest that participants with lower levels of anxiety and autism (i.e. less problematic social interaction) recruit enhanced TPJ processing to make self-serving attributions. In contrast, those with elevated levels of anxiety or autism seem to recruit less TPJ processing to make the same self-serving attributions. One potential explanation why these latter individuals easily engage in this self-serving pattern is because they immediately and automatically reject external people and situations as threatening.

Of interest also is the fact that the behavioral response does not differ between those low or high on these pathological symptoms. It may suggest that high functioning individuals with anxiety or autism may use compensatory strategies to avoid maladaptive and deviant thinking patterns that are quite successful. We therefore suggest that lowered TPJ activation during self-serving attributions might possibly serve as a neural marker of implicit clinical symptoms of anxiety and autism, which are not always revealed by self-report questionnaires. To explore this promising hypothesis, further research in sub-clinical and clinical symptoms of these psychopathologies is necessary.

CONCLUSION

The main contribution of this study is that after causal attributions to the self, the other person or the situation, we found activation in the predicted mentalizing areas responsible for causal attributions to temporary events (which excludes the mPFC), and that this pattern was

almost identical across the three attribution loci. This confirms and extends earlier findings, in particular by Kestemont *et al.* (2013), who distinguished only between attributions of another person and the situation, but did not include the self. Also new in this study are the unexpected, but robust inverse correlations between left TPJ activation and subclinical levels of anxiety and autism during external (person or situation) attributions, showing that decreased TPJ activity during self-serving attributions is associated with increased pathology. These initial results pave the way for more research on social attribution patterns in clinical populations and early clinical diagnosis in sub-clinical populations, and perhaps psychotherapeutic treatment.

REFERENCES

Arkin, R.M., Appelman, A.J., Burger, J.M. (1980). Social anxiety, self-presentation, and the self-serving bias in causal attribution. *Journal of Personality and Social Psychology*, *38*, 23–35.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17.

Blackshaw, A.J., Kinderman, P., Hare, D.J., Hatton, C. (2001). Theory of mind, causal attribution and paranoia in Asperger syndrome. *Autism*, *5*, 147–63.

Blackwood, N.J., Bentall, R.P., ffytche, D.H., Simmons, A., Murray, R.M., Howard, R.J. (2003). Self-responsibility and the self-serving bias: an fMRI investigation of causal attributions. *NeuroImage*, *20*, 1076–85.

Buchanan, G.M., Seligman, M.E.P. (1995). *Explanatory Style*. Hillsdale, NJ: Lawrence Erlbaum.

Carrington, S.J., Bailey, A.J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*, 2313–35.

Craig, J.S., Hatton, C., Craig, F.B., Bentall, R.P. (2004). Persecutory beliefs, attributions and theory of mind: comparison of patients with paranoid delusions, Asperger's syndrome and healthy controls. *Schizophrenia Research*, *69*, 29–33.

Denny, B.T., Kober, H., Wager, T.D., Ochsner, K.N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*, 1742–52.

Diez-Alegria, C., Vazques, C., Nieto-Moreno, M., Valiente, C., Fuentenebro, F. (2006). Personalizing and externalizing biases in deluded and depressed patients: are attributional biases a stable and specific characteristic of delusions? *British Journal of Clinical Psychology*, *45*, 531–44.

Dykema, J., Bergbower, K., Doctora, J.D., Peterson, C. (1996). An attribution style questionnaire for general use. *Journal of Psychoeducational Assessment*, *14*, 100–8.

Farrer, C., Frith, C.D. (2002). Experiencing oneself versus another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage*, *15*, 596–603.

Gilbert, D.T., Malone, P.S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Gilbert, D.T., Pelham, B.W., Krull, D.S. (1988). On cognitive busyness: when person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*, 733–40.

Harris, L.T., Todorov, A., Fiske, S.T. (2005). Attributions on the brain: neuro-imaging dispositional inferences, beyond theory of mind. *NeuroImage*, *28*, 763–9.

Hoekstra, R.A., Bartels, M., Cath, D.C., Boomstra, D.I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): a study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*, 1555–66.

Hope, D.A., Gansler, D.A., Heimberg, R.G. (1989). Attentional focus and causal attributions in social phobia: implications from social psychology. *Clinical Psychology Review*, *9*, 49–60.

Jones, E.E., Nisbett, R.E. (1972). The actor and the observer: divergent perceptions of the causes of behavior. In: Jones, E.E., Kanouse, D.E., Kelley, H.H., Nisbett, R.E., Valins, S., Weiner, B., editors. *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press, pp. 79–94.

Kelley, H.H. (1973). The process of causal attribution. *American Psychologist*, *28*, 107–28.

Kestemont, J., Vandekerckhove, M., Ma, N., Van Hoeck, N., Van Overwalle, F. (2013). Situation and person attributions under spontaneous and intentional instructions: an fMRI study. *Social Cognitive and Affective Neuroscience*, *8*, 481–93.

Kinderman, P., Bentall, R.P. (1996). A new measure of causal locus: the internal, personal and situational attributions questionnaire. *Person and Individual Differences*, *20*, 261–4.

Kinderman, P., Bentall, R.P. (1997). Causal attributions in paranoia and depression: internal, personal and situational attributions for negative events. *Journal of Abnormal Psychology*, *106*, 341–5.

Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*, 259–89.

Ma, M., Vandekerckhove, M., Van Hoeck, N., Van Overwalle, F. (2012). Distinct recruitment of temporo-parietal junction and medial prefrontal cortex in behavior understanding and trait identification. *Social Neuroscience*, *7*, 591–605.

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

Miller, D.T., Ross, M. (1975). Self-serving biases in the attribution of causality: fact or fiction? *Psychology Bulletin*, *82*, 213–25.

Mitchell, J.P. (2009). Social psychology as a natural kind. *Trends in Cognitive Sciences*, *13*, 246–51.

Northoff, G. (2007). Psychopathology and pathophysiology of the self in depression: neuropsychiatric hypothesis. *Journal of Affective Disorders*, *104*, 1–14.

Olson, I.R., Plotzken, A., Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, *130*, 1718–31.

Paulesu, E., Sambugaro, E., Torti, T., et al. (2010). Neural correlates of worry in generalized anxiety disorder and in normal controls: a functional MRI study. *Psychological Medicine*, *40*, 117–24.

Peterson, C., Bettes, B.A., Seligman, M.E.P. (1985). Depressive symptoms and unprompted causal attributions: content analysis. *Behaviour Research and Therapy*, *23*, 379–82.

Peterson, C., Seligman, M.E.P. (1984). Causal explanations as a risk factor for depression: theory and evidence. *Psychological Review*, *91*, 347–74.

Peterson, C., Semmel, A., von Baeyer, C., Abramson, L.Y., Metalsky, G.I., Seligman, M.E.P. (1982). The attributional style questionnaire. *Cognitive Therapy and Research*, *6*, 287–300.

Ross, L. (1977). The intuitive psychologist and his shortcomings: distortions in the attribution process. In: Berkowitz, L., editor. *Advances in Experimental Social Psychology*. New York, NY: Academic, pp. 174–221.

Ross, L.A., Olson, I.R. (2010). Social cognition and the anterior temporal lobes. *NeuroImage*, *49*, 3452–62.

Russel, D. (1982). The causal dimension scale: a measure of how individuals perceive causes. *Journal of Personality and Social Psychology*, *42*, 1137–45.

Schilbach, L., Bzdok, D., Timmermans, B., et al. (2012). Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. *PLoS One*, *7*, 1–10.

Seidel, E., Eickhoff, S.B., Kellermann, T., et al. (2010). Who is to blame? Neural correlates of causal attribution in social situations. *Social Neuroscience*, *5*, 335–50.

Seidel, E., Satterthwaite, T.D., Eickhoff, S.B., et al. (2012). Neural correlates of depressive realism: an fMRI study on causal attribution in depression. *Journal of Affective Disorder*, *138*, 268–76.

Spinhoven, P.H., Ormel, J., Sloekers, P.P.A., Kempen, G.I.J.M., Speckens, A.E.M., Van Hemert, A.M. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine*, *27*, 363–70.

Sugiura, M., Sassa, Y., Watanabe, J., et al. (2006). Cortical mechanisms of person representation: recognition of famous and personally familiar names. *NeuroImage*, *31*, 853–60.

Taylor, S., Fiske, S. (1975). Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, *35*, 439–45.

Van Overwalle, F. (1989). Structure of Freshmen's causal attributions for exam performance. *Journal of Educational Psychology*, *81*, 400–7.

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

Van Overwalle, F., Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage*, *48*, 564–84.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548–73.

Yoshimura, S., Okamoto, Y., Onoda, K., et al. (2010). Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *Journal of Affective Disorders*, *122*, 76–85.

Yoshimura, S., Okamoto, Y., Onoda, K., et al. (2013). Cognitive behavioral therapy for depression changes medial prefrontal and ventral anterior cingulate cortex activity associated with self-referential processing. *Social Cognitive and Affective Neuroscience*, 1–7.

Zigmond, A.S., Snaith, R.P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*, 361–70.

APPENDIX

Experimental sentences (best possible translation from Dutch)

List of experimental sentences
Someone helps you to learn
Someone sends you a postcard
Someone thinks that you are sensitive
Someone brings you home
Someone helps you gardening
Someone is going for a walk with you
Someone thinks that you are reliable
Someone thinks that you are interesting
Someone buys you a gift
Someone thinks that you are smart
Someone praises your new hairstyle
Someone says that he admires you
Someone invites you for a drink
Someone says that you are nice
Someone says that she respects you
Someone thinks that you are a good listener
Someone visits you to have a chat
Someone repairs your car for free
Someone thinks that you are humorous
Someone thinks that you are fair
Someone asks about your health
Someone looks forward to your visit
Someone bakes a cake for you
Someone tells you that she considers you important
Someone invites you to the cinema
Someone helps you to move out
Someone trusts you a secret
Someone says that you are reliable
Someone thinks that you are intelligent
Someone praises your tasteful clothes
Someone defends you against others
Someone thinks that you are brave
Someone thanks you for your advice
Someone greets you warmly
Someone lends you his car
Someone takes time for you
Someone appreciates your charm
Someone makes a trip with you
Someone gives water to your plants
Someone offers you to help
Someone refuses to talk to you

(continued)

Appendix (*Continued*)

List of experimental sentences
Someone thinks that you are stupid
Someone makes a hurtful comment about you
Someone starts a quarrel with you
Someone thinks that you are dishonest
Someone thinks that you are unfriendly
Someone thinks that you are unfair
Someone talks about you behind your back
Someone says that she doesn't respect you
Someone refuses to help you
Someone says that she resents you something
Someone asks you to leave
Someone ignores you
Someone doesn't show up on your birthday
Someone is disappointed in you
Someone says that he doesn't like you
Someone makes you ridiculous to others
Someone forgets an appointment with you
Someone misuses your trust
Someone says that he finds you boring
Someone doesn't return your call
Someone thinks that you are naive
Someone speaks ill of you
Someone says that you are intolerant
Someone didn't keep contact for quite a while
Someone says that you irritate him
Someone laughs at you
Someone lies to you
Someone doesn't give about your opinion
Someone doesn't visit you in the hospital
Someone lets you wait repeatedly
Someone says that she has no time for you
Someone says that he doesn't care about your problems
Someone ignores your request
Someone lets you wait for a long time
Someone doesn't invite you for her party
Someone doesn't accept your advice
Someone ignores your phone calls
Someone says that you are a coward
Someone says that your behavior is embarrassing