Virtual milgram: empathic concern or personal distress? Evidence from functional MRI and dispositional measures

Marcus Cheetham[1]*, Andreas F. Pedroni[1], Angus Antley[2], Mel Slater[2,3,4] and Lutz Jäncke[1]

[1] Department of Neuropsychology, Psychological Institute, University of Zurich, Switzerland
[2] Department of Computer Science, University College London, UK
[3] Institució Catalana de Recerca i Estudis Avançats, Universitat de Barcelona, Spain
[4] Universitat de Barcelona, Spain

One motive for behaving as the agent of another's aggression appears to be anchored in as yet unelucidated mechanisms of obedience to authority. In a recent partial replication of Milgram's obedience paradigm within an immersive virtual environment, participants administered pain to a female virtual human and observed her suffering. Whether the participants' response to the latter was more akin to other-oriented empathic concern for her well-being or to a self-oriented aversive state of personal distress in response to her distress is unclear. Using the stimuli from that study, this event-related fMRI-based study analysed brain activity during observation of the victim in pain versus not in pain. This contrast revealed activation in pre-defined brain areas known to be involved in affective processing but not in those commonly associated with affect sharing (e.g., ACC and insula). We then examined whether different dimensions of dispositional empathy predict activity within the same pre-defined brain regions: While personal distress and fantasy (i.e., tendency to transpose oneself into fictional situations and characters) predicted brain activity, empathic concern and perspective taking predicted no change in neuronal response associated with pain observation. These exploratory findings suggest that there is a distinct pattern of brain activity associated with observing the pain-related behaviour of the victim within the context of this social dilemma, that this observation evoked a self-oriented aversive state of personal distress, and that the objective "reality" of pain is of secondary importance for this response. These findings provide a starting point for experimentally more rigorous investigation of obedience.

Keywords: milgram, virtual, obedience, empathy, distress, fantasy, brain, presence

INTRODUCTION

The capacity for empathy is a fundamental aspect of effective functioning in interpersonal situations. The general view is that empathy involves the ability to share the affective state of another person (e.g., Decety and Jackson, 2004). Neuroimaging research into the mechanisms underlying affect sharing has grown in recent years, especially in connection with empathy for pain (e.g., Botvinick et al., 2005; Jackson et al., 2005). Pain-related affect sharing can lead to empathic concern for the welfare of the person experiencing discomfort and distress, and is thus regarded as an other-oriented motivated response (Batson et al., 1987; Eisenberg and Strayer, 1987; Eisenberg, 2000). The observation of another's distress can on the other hand evoke the aversive experience of personal distress. This is a self-oriented motivated response (Davis, 1980; Eisenberg and Strayer, 1987; Eisenberg, 2000) associated with fearfulness, uncertainty, emotional vulnerability, anxiety (Davis, 1996), and negative affect (Batson et al., 1987). Personal distress is posited to reflect empathic over-arousal (Eisenberg et al., 1994) and to be associated with self-regulatory effort to alleviate the aversiveness of this state (Batson et al., 1987, 1997).

Empathy is modulated by the interpersonal context in which individuals are engaged, such as competitive or co-operative interplay (Englis et al., 1982; Lanzetta and Englis, 1989; Singer et al.,

2006). One form of social interaction in which both empathy and personal distress may figure prominently is that played out in the obedience paradigm (Milgram, 1963). This paradigm induces a social dilemma in which participants feel impelled to act in compliance with the experimenter's commands to administer pain to another person while at the same time feeling driven to avoid causing harm. Slater et al. (2006) recently brought the obedience paradigm to 'virtual' life in a study using virtual reality (VR) technology. The motivation for the study was to examine the experience of presence in VR (for an overview of presence, see e.g., Sanchez-Vives and Slater, 2005). The study participants experienced discomfort and increased arousal over the course of the conflict, and some participants expressed the desire to discontinue or did actually stop applying pain to the distressed avatar. In view of the implicit assumption of Milgram's social dilemma that the participant is motivated to minimise interpersonal conflict, the implication of Slater et al.'s study is that the VR-based dilemma was transiently experienced by the participants as in some way real, and this despite the fact that the participants were left in no doubt that the avatar and her pain do not exist in the real world. It is however unclear whether the participants' affective experience during the conflict was more closely associated with empathic concern for the welfare of the avatar or with an aversive state of personal distress.

Several brain areas have been implicated in the empathic response to pain (for a review, see Decety and Jackson, 2004). These include the anterior cingulate cortex (ACC) and insula, which are associated with the affective-motivational component of the pain matrix that is also activated when a person experiences noxious stimulation directly (Morrison et al., 2004; Singer et al., 2004; Jackson et al., 2005). Changes in the hemodynamic response in these areas have been reported when participants receive pain or observe facial expressions of pain (Botvinick et al., 2005), watch videos of a stranger in pain (Morrison et al., 2004, 2007; Morrison and Downing, 2007) or view images of another person's limbs in situations that could be painful (Jackson et al., 2005). ACC and insula have also been found to correlate with the participant's judgments of the subjective severity of pain experienced by others on the basis of the other's facial pain expression (Saarela et al., 2007), and with dispositional measures of empathic concern (e.g., Singer et al., 2004).

The perception of another's emotional state is thought to automatically evoke internal emotional representations of that state (Preston and de Waal, 2002; Decety and Jackson, 2004). The involuntary nature of this process may reflect the evolutionary importance of empathic experience for human survival (e.g., Batson et al., 1987; Davis, 1996; Preston and de Waal, 2002; Decety and Jackson, 2004), even to the extent that not only the direct observation of another person's pain (Jackson et al., 2006; Singer et al., 2006) but also the perception of a signal that pain is being experienced is sufficient to induce empathy-related brain activity (Singer et al., 2004). The perception of an avatar enduring pain that in the real world would represent a fundamental challenge to survival may well have been sufficient in Slater et al.'s study to involuntarily evoke affect sharing, irrespective of the 'reality' of the pain. The immediate reality of pain may under certain circumstances be of secondary importance given that the perception of injections into an anesthetised hand known by the participants to be temporarily insensitive to painful stimulation can also induce such brain activity (Lamm et al., 2007b).

Alternatively, the participants' affective response to the avatar in Slater et al.'s study may have been rooted in an aversive state of distress as a reactive response to the pain-related distress of the avatar. The observation of another person in pain frequently triggers affective distress in the observer (Craig, 1968), and individuals are thought to be predisposed to experience another's verbal and non-verbal cues of distress as aversive (Blair, 1995). This is reflected in Davis's approach to the trait of personal distress as the "tendency to experience distress or discomfort in response to extreme distress in others" (Davis, 1994). Coping with distress involves the alleviation of its immediate aversiveness (Batson et al., 1997) and self-regulatory effort to re-establish control and goal-oriented behaviour (Sinha, 2001). But the very essence of the obedience dilemma is that participants are wedged between conflicting and in themselves adverse alternatives for which there is no prospect of establishing a personally and collectively favourable outcome.

The aim of this study was to investigate the pattern of neural responses evoked by the perception of the avatar's cues of pain-related behaviour within the context of a virtual version of the obedience paradigm. Given the complexity of psychological events in the obedience paradigm, the virtual character of the paradigm in this study, and the scant neuroimaging literature addressing affect sharing, empathy, and distress under such potentially challenging social conditions, we adopted an exploratory position with a view to providing preliminary findings for further investigation of obedience. Considering that dispositional measures of empathic concern and distress have predicted specific activity in brain regions associated with perception of another's pain (Singer et al., 2004, 2006; Saarela et al., 2007; for exceptions, see Jackson et al., 2005; Lamm et al., 2007a), we asked whether the Empathic Concern and Personal Distress subscales of the Interpersonal Reactivity Index (Davis, 1996), a well established and often used measure of different facets of empathic responsiveness, would predict changes in brain activation during the perception of the avatar's pain-related behaviour. Because of the virtual nature of our stimuli, we included the Fantasy subscale of the IRI as a measure of the tendency to imaginatively transpose oneself into fictional situations and characters (Davis, 1996).

We used functional magnetic resonance imaging (fMRI) to measure brain activity in healthy participants required to elicit virtual pain to a female avatar in a number of repeated trials within the context of a VR-based obedience paradigm. For this, Slater et al.'s virtual scenario was adapted for use in the scanner. This study focused on the period during which the participant observed the avatar's pain in response to the application by the participant of an intense electric shock. The design included pain and non-pain conditions, the former comprising trials with different levels of electric shock intensity and the corresponding behavioural response of the avatar.

MATERIALS AND METHODS
PARTICIPANTS
16 healthy right-handed adult volunteers (all female, mean age = 22.8 years; SD = 1.41) with no record of neurological or psychiatric illness were recruited for this study. Handedness was assessed with a self-rating scale (Bryden, 1970). All participants were students of the University of Zurich, native or fluent speakers of Swiss-German or High German, and naïve about the purpose of the experiment. As verified at debriefing, none of the included participants reported knowledge of the Milgram social dilemma or other aggression paradigms. Participants were informed about potential risks associated with magnetic resonance imaging and screened prior to scanning to ensure suitability for MRI. It was drawn clearly to the attention of all participants that withdrawal from the investigation was possible at any time and without having to give a reason for doing so, and participants were explicitly instructed that they would be observing a computer-generated person (avatar) and not a real person. Written informed consent was obtained before participation according to the guidelines of the Declaration of Helsinki. Each volunteer received 60 Swiss Francs for participation. The study and all procedures and consent forms were approved by the Ethics Committee of the University of Zürich.
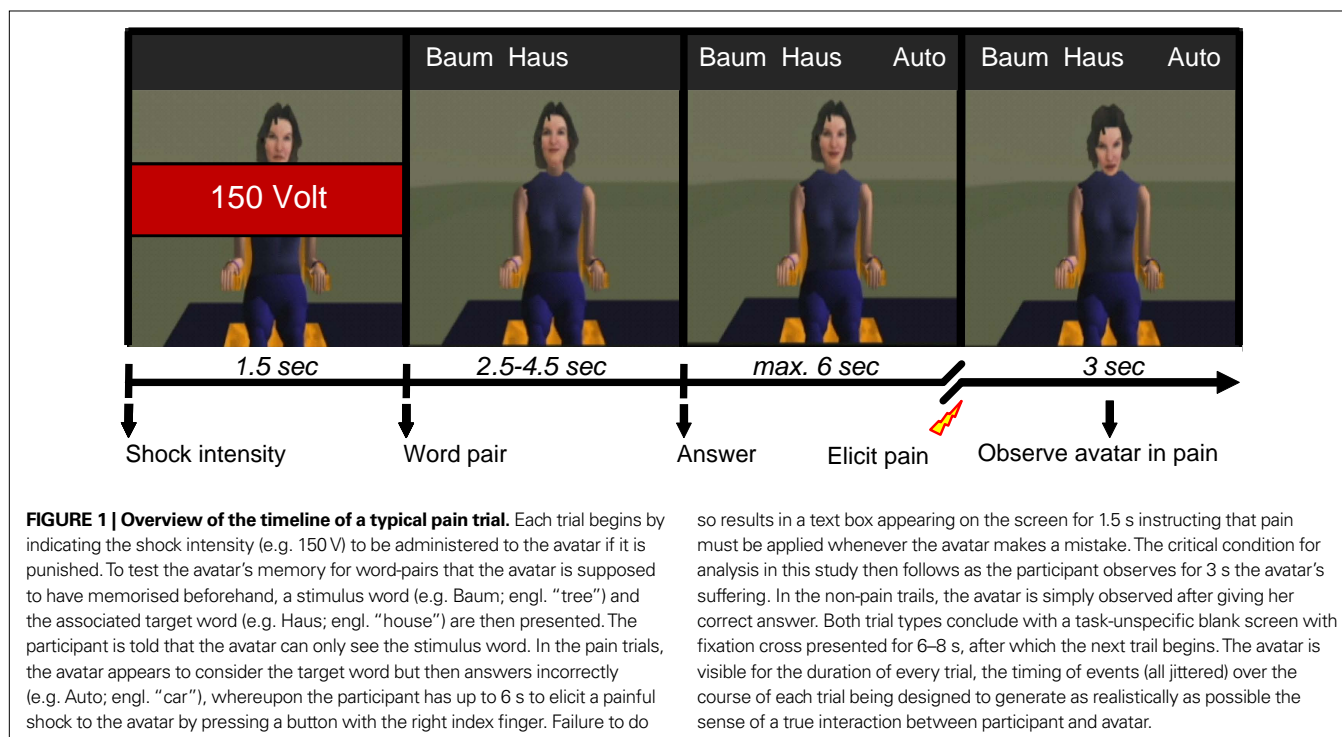
MATERIALS
Before receiving instructions, participants completed three questionnaires: (1) The Saarbrücker Persönlichkeitsfragebogen (SPF) (Paulus, 1992; German version of the Interpersonal Reactivity Index, Davis, 1996), which was administered to assess dimensions

of empathic responding. Our primary interest was to use the scores from the Empathic Concern scale (measuring prosocial feelings of warmth, compassion and concern for others) and the Personal Distress scale (measuring self-oriented feelings of anxiety and discomfort in response to the distress of others). Given the virtual nature of the stimuli and social dilemma, we included the IRI Fantasy scale (reflecting the tendency to identify with fictitious characters). The scores from the IRI Perspective Taking scale (as a measure of spontaneous adoption of the psychological point of view of others and reasoning about their mental states) were also analysed. The construct validity of the original IRI scales is reported by Davis, 1983). The internal consistencies (alpha coefficients) of the German subscales range from 0.58 to 0.73 (Paulus, 1992). (2) The NEO-FFI (Five-Factor Inventory) (Borkenau and Ostendorf, 1993; German version of the NEO-Five-Factor Inventory, Costa and McCrae, 1992) and (3) the ARES-Skalen (Hartig and Moosbrugger, 2003; based on the original BIS/BAS scales, Carver and White, 1994) were administered to gain additional insight into the predictive value of dispositional behaviour for responding in this social dilemma. The Neuroticism scale of the NEO-FFI was used to assess sensitivity to negative or punishment cues (e.g., McCrae and Costa, 1991; Costa and McCrae, 1992; Watson and Clark, 1992; Izard et al., 1993; Pickering and Gray, 2001), such as those present in negative social situations (Izard et al., 1993; Bolger and Zuckerman, 1995). The retest-stability of the German Neuroticism Scale is 0.80 (Borkenau and Ostendorf, 1993). The BIS-scale (Anxiety) of the ARES-Skalen was used to assess anxiety in terms of behavioural inhibition sensitivity. The behavioural avoidance (or inhibition) system (BIS) was originally proposed by Gray, 1982) as a behavioural regulation system responsive to aversive and punishing environmental cues. The BIS (like its counterpart, the behavioural approach system or behavioural

activation system) is suggested to be related to personality (e.g., Fowles, 1980; Gray, 1987; Gray and McNaughton, 2000). We used the BIS1 (Anxiety) scale as a measure of the vulnerability to anxiety in non-everyday situations. The internal consistency (alpha coefficient) of the German BIS1 subscale is 0.84 (Hartig and Moosbrugger, 2003).

TASK DESIGN AND PROCEDURE

Each participant was told that the task in the scanner would be to test the learning and memory performance of a computer-generated female student (avatar) in a word-pair association task for which the student had already memorised the words. In each trial, the stimulus word and the target answer of each word pair were presented in text on the screen above the avatar (see **Figure 1** for the timeline of pain trials in which the avatar should receive punishment). While both words were clearly visible to the participants, they were told that the avatar could only "see" the stimulus word and that upon seeing it should retrieve from memory and verbalise the associated target word. Because the acoustic noise of the scanner during image acquisition could impede comprehension of the avatar's response, the participants were told that avatar's answer would also be presented in text to the right of the target word and at the same time as the verbal response. For every incorrect answer, the participant was instructed to punish the avatar by pressing the appropriate button of the response box with the index finger of the right hand to deliver an electric shock. For correct answers, the participant should press the other button of the response box with the same finger. The participant was instructed that punishment must follow every incorrect answer, irrespective of how painful the electric shocks might appear to be, but not before the participant was absolutely clear in her mind that the press of the correct or incorrect button was due (i.e. that the avatar had

150 Volt	Baum　Haus	Baum　Haus　　Auto	Baum　Haus　　Auto
1.5 sec	*2.5–4.5 sec*	*max. 6 sec*	*3 sec*
Shock intensity	Word pair	Answer	Elicit pain　　Observe avatar in pain

FIGURE 1 | Overview of the timeline of a typical pain trial. Each trial begins by indicating the shock intensity (e.g. 150 V) to be administered to the avatar if it is punished. To test the avatar's memory for word-pairs that the avatar is supposed to have memorised beforehand, a stimulus word (e.g. Baum; engl. "tree") and the associated target word (e.g. Haus; engl. "house") are then presented. The participant is told that the avatar can only see the stimulus word. In the pain trials, the avatar appears to consider the target word but then answers incorrectly (e.g. Auto; engl. "car"), whereupon the participant has up to 6 s to elicit a painful shock to the avatar by pressing a button with the right index finger. Failure to do

so results in a text box appearing on the screen for 1.5 s instructing that pain must be applied whenever the avatar makes a mistake. The critical condition for analysis in this study then follows as the participant observes for 3 s the avatar's suffering. In the non-pain trails, the avatar is simply observed after giving her correct answer. Both trial types conclude with a task-unspecific blank screen with fixation cross presented for 6–8 s, after which the next trail begins. The avatar is visible for the duration of every trial, the timing of events (all jittered) over the course of each trial being designed to generate as realistically as possible the sense of a true interaction between participant and avatar.

earned the punishment or not). No instruction about the speed of the response was given but that the button must be pressed within 6 s of the avatar giving its answer. The participant was also told that the predetermined volt intensity of the electric shock to be delivered in the event of a wrong answer would be shown at the beginning of each trial (75, 150, or 225 V), and that this cue could not be seen by the avatar. After giving instructions, the participant was reminded very clearly that the pain and the avatar exist only in a virtual and not in the real world and that withdrawal from the experiment is possible at any time; once in the scanner, participants were given an alarm button to hold in the left hand throughout data registration with which termination of the experiment could be signalled.

Each scanning session consisted of three experimental runs of 25 trials each. The visual stimuli were presented using the "VisuaStim – Digital" MRI-compatible head-mounted display (Resonance Technology Inc.), using visual mono display with resolution of 800 × 600 and 30° field of view. In each run there were 7 non-pain trials and 18 pain trials, 6 for each of the three levels of intensity. The sequence of trials was pseudo-randomised under the constraint that no two trials of same shock intensity were presented in sequence. The timing of events within each trial was designed to generate as realistically as possible the sense of a true interaction between participant and avatar. All events in the trials were jittered. At the beginning of each trial (see **Figure 1**), a cue was presented for 1.5 s to indicate the predetermined intensity of the electric shock to be administered as punishment. The word pair was then presented and, after considering its response, the avatar gave either a correct or an incorrect answer (2.5–4.5 s after presenting the words). Following this, the participant had up to 6 s to press the appropriate button. After pressing the button for a correct answer the avatar was simply viewed for a further 6–8 s. If however punishment was delivered after an incorrect answer, the subsequent suffering of the avatar was observed. The suffering was conveyed by cues of pain-related behaviour (scream, facial expression of pain and corresponding movement of head posture and upper torso) at one of three possible levels of pain intensity. The stimuli for the three levels of pain intensity were originally selected on the basis that the different levels of visual and acoustic intensity were clearly distinguishable according to behavioural ratings. Failure to apply pain when punishment was due resulted in a stern reminder appearing on the screen in a text box for 1.5 s duration, demanding that pain be applied whenever the avatar makes a mistake. After observing the avatar in pain for 3 s, the avatar was viewed for a further 3–5 s during which time the avatar showed no pain-related behaviour. To ensure that the fMRI signal could return to a task-unspecific baseline, a blank screen with a fixation cross was subsequently presented for both trial types for 6–8 s after which the next trial began. Trials in which participants failed to respond appropriately were discarded in the data analyses. The total scanning time was approximately 35 min. Participants were debriefed after scanning about their experience during the social dilemma with particular reference to the impact of observing the avatar in pain.

fMRI DATA ACQUISITION AND ANALYSIS OF IMAGING DATA
Structural and functional images were acquired with a 3-T whole-body MR unit (Philips Medical Systems, Best, The Netherlands) and eight-channel Philips SENSE head coil. Structural images of the entire brain using a T1-weighted three-dimensional, spoiled gradient echo pulse sequence (180 slices, TR = 20 ms, TE = 2.3 ms, flip angle = 20°, FOV = 220 mm × 220 mm × 135 mm, matrix size = 224 × 187, voxel size = 0.98 mm × 1.18 mm × 0.75 mm, resliced to 0.86 mm × 0.86 mm × 0.75 mm). Functional images were acquired from 225 whole-head scans per run using a Sensitivity Encoded (SENSE) (Pruessmann et al., 1999) single-shot echo-planar imaging technique (repetition time, TR = 2.5 s; echo time, TE = 35 ms; field of view = 220 mm × 220 mm × 132 mm; flip angle = 78°; matrix size = 80 × 80; voxel size = 2.75 mm × 2.75 mm × 4 mm, resliced to 1.72 mm × 1.72 mm × 4 mm). Three dummy scans at the beginning of each run were acquired and discarded in order to establish a steady state in T1 relaxation for all functional scans.

fMRI DATA ANALYSIS
MATLAB 2006b (Mathworks Inc., Natick, MA, USA), and the SPM5 software package (http://fil.ion.ucl.ac.uk/spm) were used for pre-processing and MRI data analysis. All images were realigned to the first recorded volume, normalised into standard stereotactical space (using the EPI-template provided by the Montreal Neurological Institute, MNI brain), resliced to 2 mm × 2 mm × 2 mm voxel size and smoothed using a 6-mm full-width-at-half-maximum Gaussian kernel. Activated voxels were identified by a general linear model (Friston et al., 1995) implemented in SPM5. First, a statistical model for each subject was computed and high-pass filtering applied (cut-off 128 s). We separately modelled the "pain" and "non-pain" trials, modelling the epoch as beginning with the button press and lasting in both trial types for 3 s. The design matrix included also regressors for the initial cue at the beginning of each trial indicating the intensity of the shock that participants would expect to have to elicit. All regressors were convolved with SPM's canonical hemodynamic response function.

To investigate the contrast of interest, that is, brain areas showing regionally specific effects of observing the avatar in pain, the estimates derived from the first level (fixed-effects) analysis of the pain condition were compared with the non-pain condition, thus producing a statistical parametric map of the t statistic SPM {t} of values for each voxel with which to test for the condition effects. For this, linear contrasts were applied for each voxel for each subject (Friston et al., 1995), this resulting in one statistical parametric map for the contrast for each subject. We were further interested in identifying brain regions that show a monotonic dependency in hemodynamic response over the three different levels of pain trail types (i.e., 75, 150, or 225 V) and their corresponding intensity of pain-related behaviour and affect expression of the avatar. Additional contrasts were therefore employed, modelling a linear dependency between hemodynamic responses and the three levels of pain. All contrast images were smoothed using a 8-mm FWHM Gaussian kernel. Second-level analysis was performed to compare activation values between the pain and non-pain conditions using one-sample random effects t-statistics across subjects to allow for population inferences.

Guided by previous neuroimaging studies of significant changes in the hemodynamic response to pain experience, dynamic facial expressions of pain, and empathy for pain (e.g., Simon et al., 2006; Lamm et al., 2007a), we defined for the observation of the avatar

in pain *a priori* regions of interest (ROI) that included amygdala, ACC, Insula, somatosensory cortices (SI and SII/posterior insula), inferior frontal cortex (IFG), supplementary motor area (SMA), premotor cortex (PMC), and fusiform gyrus. To explore a broad range of effects in the data, voxels are reported that survived significance thresholding at $p < 0.001$, uncorrected for multiple comparisons with a spatial extent threshold of $k = 10$ voxel. Activations found in these regions of interest are reported in **Table 1**. The used statistical threshold in this contrast is higher and thus more conservative than in other studies exploring emotional brain responses (e.g., Hamann and Mao, 2002).

To explore whether brain activity during observation of the avatar in pain covaried with individual differences in behavioural data of dispositional Empathic Concern, Personal Distress, and Fantasy (as well as Perspective Taking, Neuroticism, and Anxiety), the scores from the corresponding measures were entered separately in the GLM as covariates in the second level random effects analysis. To ensure that disposition-related differences in neural activity were orthogonal to the pain > non-pain contrast and to avoid a large quantity of false positives, the same *a priori* defined regions of interest as used for the pain > non-pain contrast were applied in this analysis. Given the "make-believe" nature of the pain and suffering in this study, we defined also *a priori* regions associated in a previous study (Jackson et al., 2006) with imagined pain in an artificial limb. These regions were medial prefrontal cortex (MPFC),

posterior cingulate cortex, and inferior parietal lobule, and we report activations within these ROIs also for the pain < non-pain contrast. Voxels are reported that survived significance thresholding at $p < 0.001$, uncorrected for multiple comparisons with a spatial extent threshold of $k = 10$ voxel. In **Table 2**, only activations found in these *a priori* regions are reported. Activations falling outside these ROIs are not reported or interpreted.

RESULTS
BEHAVIOURAL DATA
To test whether participants responded differently in this dilemma as a function of shock intensity, a mixed-effects ANOVA was performed. Greenhouse-Geisser adjustment was applied to correct the degrees of freedom for violation of the assumption of sphericity. The analysis revealed a significant interaction between condition (volt level: 0, 75, 150, and 225) and reaction time (RT) [$F(1.36, 20.42) = 24.9$, $p = < 0.001$]. The tests of within-subject contrasts showed a significant difference in RT between pressing the response button that elicited no shock and pressing the button that elicited a shock of 75 V [$F(1,15) = 31.71$ $p = < 0.001$], 150 V [$F(1,15) = 25.79$ $p = < 0.001$], or 225 V [$F(1,15) = 28.71$ $p = < 0.001$], such that the RT was shorter for the non-pain condition in which no shock was elicited. There were no significant differences between the different levels of shock (see **Figure 2**). All but two of the participants pressed the shock button correctly in all trials. One participant failed to

Table 1 | Regions of significant activation during observation of the avatar in pain compared with no pain in *a priori* defined regions of interest.

Coordinates in MNI space and maximum t values are shown for local voxel maxima in each cluster (SPM{t} maps thresholded at $p < 0.001$ uncorrected for multiple comparisons, with a cluster extent threshold of 10 voxels).

Region of activation	BA	L/R/M	x	y	z	Max *t* value	Cluster size
Superior frontal gyrus	10	R	12	56	30	5.92	370
º Superior frontal gyrus	9/10	R	18	52	42	5.73	
º Superior frontal gyrus	10	M	0	56	28	5.05	
Superior frontal gyrus	8	L	−10	48	48	7.19	259
º Superior frontal gyrus	8	L	−20	42	48	4.04	
Precentral gyrus	4	R	42	−10	52	4.71	142
º Dorsal premotor cortex	6	R	38	−14	44	4.25	
º Precentral gyrus	4	R	46	−10	60	4.17	
Dorsal premotor cortex	6	R	20	−12	74	4.20	11
Uncus/temporal pole of STG	28/38	R	32	6	−24	5.67	269
º Amygdala		R	16	−4	−20	3.98	
Posterior cingulate	31	R	18	−64	−8	4.08	426
º Parahippocampal gyrus	30/27	L	−12	−40	−8	3.92	
º Parahippocampal gyrus	30	L	−16	−50	−2	3.39	
Inferior frontal gyrus (oper.)	44	R	50	44	2	5.01	97
º Inferior frontal gyrus (orb.)	47	R	54	36	−2	4.42	
º Inferior frontal gyrus (orb.)	47	R	52	38	−12	3.98	
Inferior frontal gyrus (orb.)	47	R	34	30	−18	5.38	67
º Inferior frontal gyrus (orb.)	47	R	42	26	−16	4.81	
Inferior frontal gyrus (orb.)	47	L	−48	24	−2	8.11	243
Fusiform gyrus	19/37	L	−46	−54	−26	5.51	351
º Fusiform gyrus	18/19	L	−38	−74	20	4.98	
º Cerebellum, decline		L	−40	−66	22	4.95	

º, subpeaks of a cluster; BA, approximate Brodmann area; L, left hemisphere; R, right hemisphere; M, medial activation; STG, superior temporal gyrus; MTG, medial temporal gyrus; oper., opercularis; orb., orbitalis.

Table 2 | Sites of activation related to Distress, Fantasy, Neuroticism and Anxiety in *a priori* defined regions of interest. Coordinates in MNI space and maximum t values are shown for local voxel maxima in each cluster (SPM{t} maps thresholded at *p* < 0.001 uncorrected for multiple comparisons, with a cluster extent threshold of 10 voxels).

Region of activation	BA	L/R/M	*x*	*y*	*z*	Max *t* score	Cluster size
DISTRESS							
Superior frontal gyrus	10	R	14	52	26	5.45	34
Inferior frontal gyrus (tri.)	45	R	42	30	14	5.73	332
° Inferior frontal gyrus (orb.)	47	R	52	34	−2	4.66	
Inferior frontal gyrus (orb.)	47	R	28	32	−14	5.39	21
Uncus	28	L	−16	−10	−16	5.26	91
° Amygdala		L	−18	−2	−12	4.43	12
FANTASY							
Superior frontal gyrus	9/10	R	16	46	38	4.35	19
SMA	6	L	−8	−14	54	4.82	27
NEUROTICISM							
Superior frontal gyrus	10	R	14	48	36	5.16	19
Inferior frontal gyrus (oper.)	44	R	48	42	4	4.75	38
ACC	32	M	−2	46	10	4.99	80
ACC	32	M	0	48	12	4.22	20
ACC	24	M	−6	24	26	4.14	12
Anterior VMPFC	10	M	2	54	−6	4.19	11
Middle frontal gyrus	9	L	−20	38	42	4.08	10
SMA	6	R	8	−12	56	6.94	312
° SMA	6	R	6	−4	66	5.52	
° SMA	6	M	−2	−12	70	4.60	
BIS 1 (ANXIETY)							
Inferior frontal gyrus (orb.)	47	L	−54	24	12	7.19	241
° Middle frontal gyrus	9	L	−50	16	30	6.73	
Inferior frontal gyrus (oper.)	44	R	62	12	16	4.46	34
Inferior frontal gyrus (orb.)	47	R	54	30	0	6.11	28
Amygdala		L	−24	2	−16	4.34	24

tri., triangularis; VMPFC, ventromedial prefrontal cortex. Refer to **Table 1** for other abbreviations.

elicit pain of 150 and 250 V on three trails in the first and second run, and the other did not trigger five 150 and five 250 V shocks in the first run.

IMAGING DATA

Network of areas involved in the observation of the avatar in pain
The first analysis sought to establish the pattern of activation associated with the perception of the avatar's cues of pain-related bodily and facial expression and pain vocalisation. For this, the pain > non-pain contrast revealed activations within a number of the *a priori* regions of interest, as shown in **Table 1**.

Predominantly right-sided activity was found in areas associated with detecting and evaluating affective information. This included amygdala (see **Figure 3A**) within a cluster that also encompassed the uncus and temporal pole of the superior temporal gyrus. There was also activation in ventrolateral parts of inferior frontal gyrus (IFG) (see **Figure 3B**), most prominently right and left opercular IGF (BA47) and right orbital IFG (BA44), and activation in posterior cingulate in a cluster including retrosplenial cortex and posterior parahippocampus. Activation of motor

areas was limited to the dorsal premotor cortex (PMC) (BA6) and precentral gyrus (BA4). The analysis showed activation in BA8 of the superior frontal gyrus and in the rostral prefrontal cortex (BA10) (see **Figure 3C**). There was also activation in the fusiform gyrus. No activity was found in the pain > non-pain contrast in areas often associated with affect sharing such as the ACC, insula, supplementary motor area (SMA) or pre-SMA, ventral PMC or Broca's area in the left IFG, even at the very liberal threshold of *p* > 0.05. No differences in activation were found as a function of applied shock intensity.

Dispositional measures and brain activity during observation of avatar in pain
Changes in neuronal activity during perception of the avatar in pain covaried linearly with the scores from the Distress, Neuroticism, Anxiety, and Fantasy scales, such that higher trait scores were associated with increasing neuronal activity in the areas shown in **Table 2**. The scores on the Empathic Concern and Perspective Taking scales of the IRI failed to predict any single voxel of change in neuronal activity, even at the very liberal threshold of *p* < 0.05.
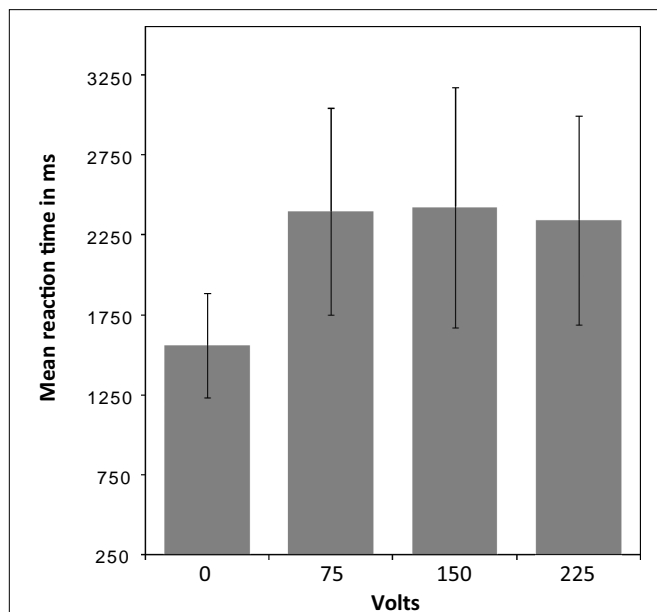
FIGURE 2 | Mean reaction time for pain and non-pain button responses.
A mixed-effects ANOVA (Greenhouse-Geisser adjusted) revealed a significant interaction between condition [volt level: 0 (i.e., non-pain condition), 75, 150, and 225] and reaction time (RT) [$F(1.36,20.42) = 24.9$, $p = < 0.001$]. Within-subject contrasts showed a significant difference in RT between the non-pain condition and the pain condition as a whole, with shorter RT for the non-pain condition, but no significant differences between the shock levels within the pain condition.

FIGURE 3 | Significant hemodynamic activity during observation of avatar in pain compared with not in pain. Observing the avatar in pain compared with no pain evoked differences in brain activation in the right amygdala and periamygdala areas **(A)**, IFG **(B)**, and bilateral rostral PFC (rPFC) **(C)**. Images are superimposed on the coronal and sagittal sections of the single subject structural MNI template.

Distress, Neuroticism, and Fantasy (but not Anxiety) (see **Figures 4B–D**) covaried with activity in the rostral PFC (BA10) that spatially falls within the cluster of activity found in the same region in the pain > non-pain contrast (see **Figure 4A**). Fantasy did not covary with any other cluster identified in the pain > non-pain contrast, but it did covary with SMA. Neuroticism also covaried with SMA and ACC (see **Figure 5A**). Distress (see **Figure 5B**) and Anxiety were related to left amygdala. Distress was also associated with activation in right orbital (BA47) and triangular (BA45) IFG, whereas Anxiety predicted activation in left and right orbital (BA47) (see **Figure 5C**) and right opercular (BA44) IFG.

DISCUSSION

This exploratory study took a simple approach to gaining first insight into the complex interplay of social and emotional processing in a VR-based version of Milgram's obedience dilemma. For this, we adapted as faithfully as possible the virtual stimuli of Slater et al.'s study for use within the constraints of fMRI methodology. Given the obvious virtual nature of the interpersonal setting, the victim and her pain, we asked whether observation of the avatar in pain does evoke brain activity consistent with an aversive state of personal distress or with empathic concern for the avatar. The results show that the perception of the avatar's combined pain-related bodily and facial expression and pain vocalisation evoked neuronal activity in a distributed network of brain areas known to be involved in affective processing, but not with those areas commonly reported in other studies of pain-related affect sharing and empathy. Supported by the exploration of whether and
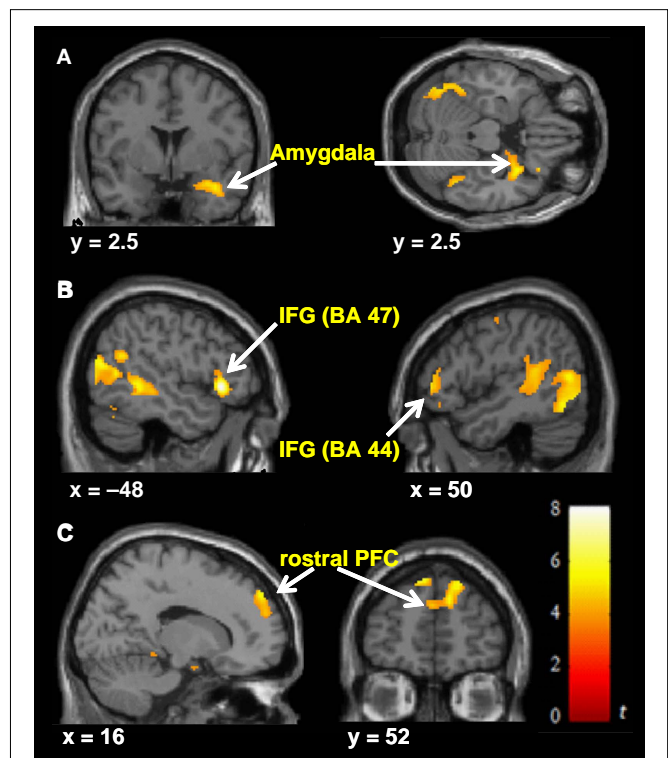
which of these regions were modulated by dispositional behavioural tendencies, the overall findings are consistent with the idea that this paradigm induced an aversive state of personal distress and not empathic concern.

Davis (1996) put forward the view that personal distress may be a simple reactive response to another's display of extreme distress without representing the other's state. A fast, involuntary affective response was most probably elicited by the observation of the avatar's audio-visual display of pain-related behaviour and expressed affect, regardless of whether the avatar was experienced as actually having an emotional state. Several studies have reported activation of structures such as the amygdala by negatively valenced stimuli, including images of human injury and graphic violence, but without activation consistent with empathy-related processing (e.g., Hamann and Mao, 2002; Schienle et al., 2006). At debriefing, most participants described the emotional display of the avatar as conveying anger, and two reported having perceived overtones of fear in her pain-related affect. Anger and fear are understood as distress emotions (see e.g., Berenbaum et al., 1995, 2003), and, importantly, the morphology of the facial expression of acute pain is very similar to that of anger and to a lesser degree fear (see e.g., Williams, 2002). In fact, the pattern of brain activations found in this study does mirror in part that reported by Pichon et al. (2008) in their study of dynamic bodily expression
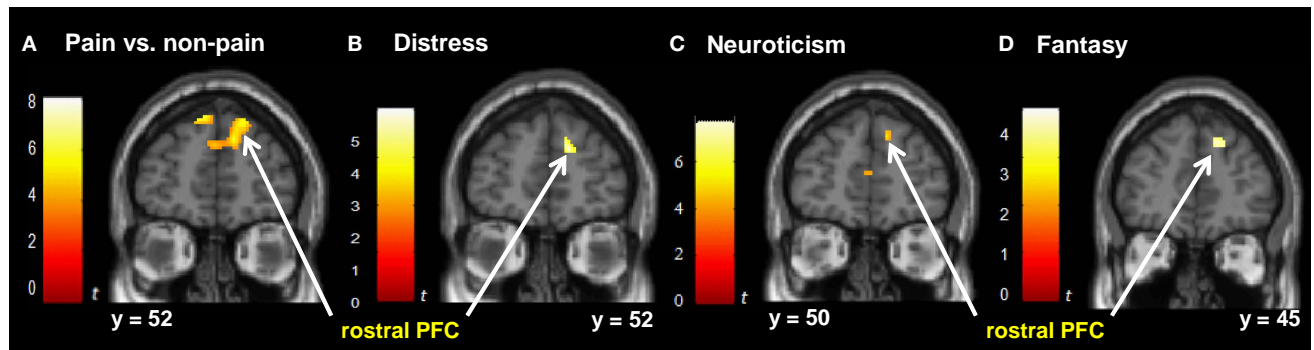
FIGURE 4 | Correlated dispositional measures and rostral PFC activity during observation of avatar in pain. The activation of the rostral PFC in the pain vs. non-pain contrast is shown in **(A)**, and, as found in the subsequent analysis, the prediction by scores of Personal Distress **(B)**, Neuroticism **(C)**, and Fantasy **(D)** of change in hemodynamic response in rostral PFC during observation of the avatar in pain, superimposed on the coronal sections of the single subject structural MNI template.
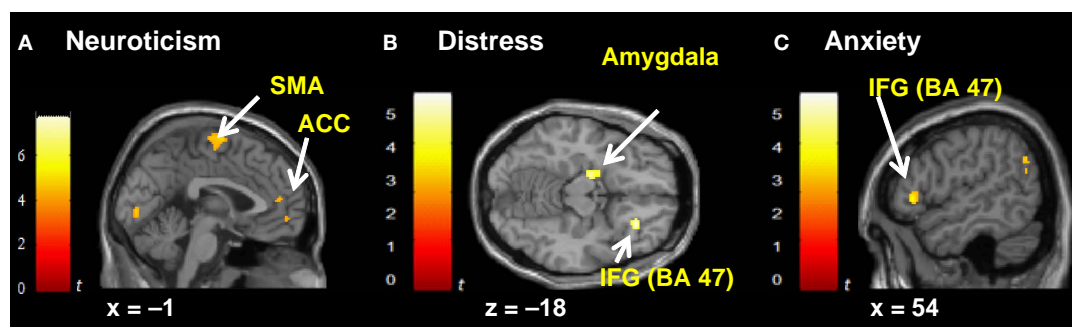
FIGURE 5 | Correlated dispositional measures and brain activity during observation of avatar in pain. Prediction of hemodynamic changes in activation during observation of the avatar in pain versus no pain by scores of Neuroticism **(A)**, Distress **(B)** and Anxiety **(C)**, superimposed on the coronal and sagittal sections of the single subject structural MNI template.

of anger. Interestingly, their study found no differential activity in ACC and insula activity in the comparison of anger and neutral expression. Irrespective of these considerations, it is important to note that a simple reactive response to the negatively valenced display of the avatar may well have occurred independently of the other events in the paradigm and of those mechanisms of obedience that this paradigm is intended to induce. But, given that the anticipation and subsequent perception of the victim in pain is a vital component of Milgram's social dilemma, the nature of the participant's response to pain-related behaviour in this paradigm is important for understanding obedience.

It is possible that a shared representation of affect was elicited by the perception of the avatar and her apparent pain. This would assume that the avatar's behaviour was represented and processed as if communicating an internal affective state. This is plausible given that the audio-visual stimuli were carefully designed to vividly and dynamically convey pain-related behaviour and spontaneous affect in a real-time interactive manner as consistent as possible with human experience, and that affect sharing may be induced even though the participant knows that the target of noxious stimulation can actually feel nothing (Lamm et al., 2007b). Assuming that processes of affect sharing are engaged, Decety and Jackson

(2004) propose that it is the failure to apply sufficient self-regulatory emotional control over the shared state that leads to the experience of emotional distress or emotional contagion. The latter is an affective state that matches the other's emotional display (Decety and Meyer, 2008), and de Waal (1996) proposes that cognitive failure to correctly attribute shared affect to the eliciting source (i.e., no self-other distinction) can lead to emotional contagion. The cognitive components of empathy, which are thought necessary for the experience of empathic concern, are suggested to include the abilities to correctly attribute and effectively regulate shared affect as well as to adopt the perspective of the other (Decety and Jackson, 2004). But our measures of empathic concern and perspective taking did not predict any brain activation. Taken together, and while keeping in mind the exploratory nature of the findings, the makeup of the neural network reported in this study may reflect the basic pattern of brain activation underlying self-oriented distress upon witnessing the pain of one's victim within the context of the obedience paradigm. This pattern has a number of main features.

Observing the avatar in pain elicited greater activation in the right amygdala. The amygdala is associated with feelings of fear and anxiety in response to electrical stimulation (Halgren et al., 1978), with the processing of facially (Adolphs et al., 1994; Breiter

et al., 1996; Calder et al., 1996) and vocally conveyed expression of threat, fear and anger (Scott et al., 1997; Phillips et al., 1998; Isenberg et al., 1999), and with dynamic bodily expression of fear (Grezes et al., 2007) and anger (de Gelder et al., 2004). In contrast to our finding of amygdala activation, Simon et al. (2006) report amygdala deactivation during viewing of dynamic pain expression of natural female faces. They suggest that this may reflect the suppression of avoidance and the mobilisation of helping behaviour, which in turn is an important precondition for empathic concern (Decety and Moriguchi, 2007). And, Petrovic et al. (2004) propose that deactivation in the amygdala during direct pain (e.g., Petrovic et al., 1999) may stem from top-down regulation of aversive personal distress in a "no-escape" situation. These suggestions echo the established importance of the amygdala as a target of adaptive voluntary regulation of negative affect (e.g., Schaefer et al., 2002; Ochsner et al., 2004; Beauregard et al., 2006), for which anterior cingulate cortex (ACC) and ventrolateral prefrontal cortex (VLPFC) (BA47) play a prominent role. For instance, successful cognitive reinterpretation of threat-related stimuli as a means to attenuating negative affect is associated with deactivation of the right amygdala and increased activation in the ventral and dorsal prefrontal cortex, including ACC (Ochsner et al., 2002; see also Eippert et al., 2007). Another such strategy entails cognitively evaluating the level of realism of fearful and threatening stimuli as either artificial or natural, this also resulting in attenuation of right amygdala activation and an increase in the response of the ACC and right prefrontal cortex, including BA47 (Hariri et al., 2003). The impact on empathic experience of consciously reappraising the context in which pain-related stimuli are perceived is well documented (e.g., de Vignemont and Singer, 2006; Jackson et al., 2006; Singer et al., 2006; Lamm et al., 2007a). While our findings show that observation of the avatar's pain-related behaviour evoked activation in right and left orbital (BA47) and right opercular (BA44) VLPFC, there was no evidence of any ACC activation in the comparison of the pain and non-pain conditions.

A modulatory role for affective processing has been implicated in right VLPFC in association also with regulating or suppressing pain-related negative affect and distress (Petrovic and Ingvar, 2002; Wager et al., 2004; Lieberman et al., 2007), and with actively ignoring angry voices (Sander and Scheich, 2001). In fact, right amygdala and BA47 appear to be functionally associated in responding to angry and fearful faces such that the activation of BA47 is conversely related to that of the amygdala (Hariri et al., 2000, 2003; Nomura et al., 2004). Nomura et al. (2004) found that this region is associated with conscious evaluation of affective expression. The right BA47 is sensitive to both the observation and experience of pain (Lamm et al., 2007a), to static pictures of anger expression (Blair and Curran, 1999), dynamic fearful body expression (Grezes et al., 2007), angry and fearful tone of voice (Wildgruber et al., 2005) – which also activates right opercular BA44 – and nonlinguistic vocalisations of female cries and fearful screams (Fecteau et al., 2005). While participants reported great difficulty ignoring the painful cries of the avatar, the actual nature of the functional balance between BA47 and amygdala in this study is unclear. Activation of this right orbital area was predicted by dispositional distress.

The ACC and insula have been linked with the affective and motivational components of processing another's perceived pain (Morrison et al., 2004; Singer et al., 2004, 2006; Botvinick et al., 2005; Jackson et al., 2005, 2006). But we found no differential activation of these structures in the pain and non-pain comparison. It is possible that this paradigm evoked an unspecific aversive state of arousal, possibly associated with distress and anxiety (Eisenberg, 2000; Critchley et al., 2004) and a state of avoidance (Decety et al., 2007), across the pain and non-pain conditions, and that this state was related also to other events and subjective experiences within the paradigm and not just the observation of pain. In keeping with the preceding considerations, one interpretation would be that ACC and insula were not differentially engaged in processing and regulating negative affect specifically in response to observing the avatar in pain compared with her not in pain. Failure to down-regulate arousal can lead to distress (Decety and Moriguchi, 2007), associated overarousal (Eisenberg et al., 1994), and feelings of fearfulness and anxiety (Davis, 1996). In contrast, top-down regulation of arousal elicited by the observation of the affective state of another is thought necessary to experience empathy (Eisenberg, 2000; Decety and Jackson, 2006; Decety and Moriguchi, 2007). Individual differences in neuroticism did however predict change in brain activation of the ACC during observation of pain, this possibly reflecting trait-consistent differences in the experience of negative affect and maladaptive coping and regulation of affect during exposure to cues of threat and punishment; this awaits further investigation.

The perception of the avatar in pain also evoked activation in the left orbital (BA47) VLPFC. This area is responsive to observation of oral articulatory movements (Buccino et al., 2004) and voiced speech production (Schulz et al., 2005), and, like its right homologue, to non-linguistic vocalisations such as cries and fearful screams (Fecteau et al., 2005) and to observation of fearful body expressions (de Gelder et al., 2004). Saarela et al. (2007) found activation in this region in response to the presentation of faces expressing intense pain. This area is responsive also to negatively valenced words (e.g., suicide, victim) (Engels et al., 2007) and is involved in giving meaning to linguistic speech (Petersen et al., 1990; Dapretto and Bookheimer, 1999). It is possible that this region's activation was related in this study to giving meaning to the alarming pain-related visual and auditory cues. The neighbouring left areas of Broca (BA44) and premotor cortex (BA6) are associated with the role of the IFG in observing, imitating and comprehending action and are underpinned by the system of motor mirror neurons (e.g., Rizzolatti et al., 2001; Rizzolatti and Craighero, 2004) that, in turn, is thought by some to facilitate affect sharing by automatically simulating the affective behaviour of others (Decety and Jackson, 2004; Gallese et al., 2004). We found no activation in these regions of action representation in the pain vs. non-pain contrast.

While the orbital part of left VLPFC was associated with the measure of personal distress in Saarela et al.'s (2007) study, we found an association with trait anxiety in this region. Anxiety (and distress) also predicted activation in the left amygdala. The anxiety scale was used in this study to assess self-reported experience of behavioural inhibition in response to environmental punishment cues. The behavioural inhibition system (BIS) is thought to be mobilised when active avoidance of threat is perceived as ineffective or not possible in a perceived no-escape situation (Gray, 1982). Activation of the amygdala in the context of pain, distress

and non-escape has been suggested by Petrovic et al. (2004). In fact, the obedience paradigm is suggested to induce the subjective perception of diminished personal control (Milgram, 1974). In this connection, activation of the SMA might be a point of interest. It has been suggested that fear-induced contagion is allied with the preparation of action in response to the perception of another person's threatening situation and that this preparatory response principally recruits the supplementary motor area (SMA) and pre-SMA (de Gelder et al., 2004). This idea finds support in that the SMA is implicated in linking unpleasant emotions with motor action in response to aversive stimuli (Oliveri et al., 2003) and that the pre-SMA is responsive to the arousal levels of affective signals (Warren et al., 2006). We found no pre-SMA or SMA activation in the pain vs. non-pain contrast. This may be attributable to the situation being perceived as inescapable and to processes of behavioural inhibition: The BIS responds to signals of fear and danger by inhibiting motor behaviour, increasing arousal and enhancing attention to threat. But neuroticism and fantasy were associated with changes in activation of the SMA, suggesting at least that any activation in this region during observation of pain was modulated to some degree by individual differences in these dispositions.

We suggest that our findings are consistent with the idea that this paradigm induced an aversive state of personal distress and not one of empathic concern. But distress and empathy could conceivably occur concurrently in the VR-based obedience paradigm, meaning that the suggested absence of empathy needs to be addressed. All but two of the participants complied fully with the requirement to administer pain to the avatar. It is therefore possible that the participants generally submitted to the perceived authority of the experimenter and did so for example in way consistent with Milgram's (1974) proposed agentic shift of responsibility. The relinquishment of personal responsibility for the welfare of the victim may lead to a down-regulation of empathic experience. One potential reason for this high level of compliance is that the experimenter in the paradigm, one of the authors, may have been perceived by the participants as highly persuasive in his role as the expert or authority. If so, this would highlight also the importance of the 'real' components of the VR-based dilemma in augmenting obedience. Another possibility is that the participants may have identified with their role as punisher, effectively sanctioning the active use of excessive aggression. But the individual scores of the NEO-FFI personality dimension 'social agreeableness' indicate that all participants report the tendency to use non hostile-aggressive means to resolving social conflict. That the measure of empathic concern did not predict any brain activation may simply stem from the explicit instruction before entering the scanner that the pain and the avatar existed only in a computer-generated world. Once in the scanner, this information was underscored throughout the course of the experiment by the obvious virtual nature of the visual stimuli, and the participants may have proceeded through the task without losing sight of the fact that the avatar knows in reality no pain. But explicit knowledge that another person is transiently insensitive to painful stimulation is in itself not enough to inhibit brain activity associated with pain-related affect sharing and empathy (Lamm et al., 2007b), and very unrealistic cartoon-based depictions of affective situations have been used to demonstrate brain activity associated with empathy and theory of mind processing

(Völlm et al., 2006). On the other hand, awareness of the realism of painful situations depicted in the form of real or make-believe cartoon images is associated with a relative attenuation of brain activity that is, nevertheless, thought to be empathy-related (Gu and Han, 2007).

Importantly, this VR-based paradigm is understood as a "playful" rehearsal of a potentially real, novel, but dangerous situation without the participant having to incur the cost of causing harm or life-threatening injury (for this view on play, see e.g., Aldis, 1975). As playful rehearsal, the mechanisms underlying obedience and the response to viewing the avatar in pain may not therefore be fully engaged or may be altered in some way. But in playful rehearsal there is the distinction between 'role play' as a form of pretence and play as non-pretence in which the game and the player's role is experienced as if for real (Steen and Owens, 2001). The fantasy scale predicted neural modulatory brain activity during observation of the avatar in pain. This scale measures the tendency to transpose oneself into fictional situations and the feelings of characters (Davis and Franzoi, 1991), and to imagine the potential outcome of such situations (Davis, 1983). Slater et al.'s study was related to the interest in presence. Presence may be understood as the subjective experience of feeling transiently unaware of actions and cognitions linked to the real environment while perceiving oneself as situated in and actually responding to events mediated by the virtual environment as if they were really happening (Wirth et al., 2007). The experience of fantasy and presence appear to require the capacity to selectively attend to current sensory input and to concomitantly suppress attending to internal or external stimuli that could conflict with or trigger disengagement from that input (see also Slater, 2002; for further reading and findings on neural mechanisms of presence, see Baumgartner et al., 2008; Jäncke et al., 2009), including cognitions concerning the medium by which this information is actually delivered to the senses and the knowledge that the perceived events are not real. This is consistent with the idea of plausibility in presence research. Plausibility refers to the illusion that the perceived events in the virtual world are really happening, even though the participant knows that they are not (Slater, 2009). This implies a sense of "informed consent" on the part of participants in engaging to a lesser or greater extent in this illusion.

The intention behind this study was to provide preliminary findings and points of discussion to promote further neuroimaging studies of behaviour in the obedience paradigm. Our simple approach to examining brain behaviour enabled us to capture a sense of the spatially distributed response of brain regions during observation of the avatar within the context of this partial replication of the obedient paradigm. How well the neural processes engaged in a true re-run of the original Milgram paradigm correspond with those of our VR-based findings may never be established. Our approach focused on female participants only, and it was taken at the cost of controlling the many complex aspects of emotional and cognitive processing along the entire timeline of stimulus presentation and of controlling the impact of these on the time window of pain observation. Within this window itself, the differential influence of the specific elements of the visual and auditory pain-related display under conditions of obedient compliance await further investigation.

The findings of this exploratory study are consistent with the idea that the VR-based obedience paradigm evoked a self-oriented state of personal distress in response to observing the avatar in pain. This distress may be rooted in a simple involuntary response to the avatar's pain behaviour and expression of negatively valenced affect, or it may reflect poor self-regulatory control of the shared affect evoked by the perceived emotional state of the avatar. The results are consistent with the suggestion that the objective 'reality' of pain is of secondary importance in the experience of distress, and that a particular pattern of brain networks may underpin this experience. The many aspects of this social dilemma are still to be elucidated, but in throwing some light on the neural basis of one aspect of this dilemma and associated affective experience a starting point for more rigorous investigations has been established.

ACKNOWLEDGMENTS

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372, 669–672.

Aldis, O. F. (1975). Play Fighting. New York, Academic Press.

Batson, C. D., Fultz, J., and Schoenrade, P. A. (1987). Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. *J. Pers.* 55, 19–39.

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein T. R., and Highberger, L. (1997). Empathy and attitudes: can feeling for a member of a stigmatized group improve feelings toward the group? *J. Pers. Soc. Psychol.* 72, 105–118.

Baumgartner, T., Speck, D., Wettstein, D., Masnari, O., Beeli, G., and Jancke, L. (2008). Feeling present in arousing virtual reality worlds: prefrontal brain regions differentially orchestrate presence experience in adults and children. *Front. Hum. Neurosci.* 2, 8. doi: 10.3389/neuro.09.008.2008.

Beauregard, M., Paquette, V., and Levesque, J. (2006). Dysfunction in the neural circuitry of emotional self-regulation in major depressive disorder. *Neuroreport* 17, 843–846.

Berenbaum, H., Fujita, F., and Pfennig, J. (1995). Consistency, specificity and correlates of negative emotions. *J. Pers. Soc. Psychol.* 68, 342–352.

Berenbaum, H., Raghavan, C., Le, H., Vernon, L., and Gomez, J. (2003). Disturbances in Emotion. New York, Russell Sage Foundation.

Blair, R. J. (1995). A cognitive developmental approach to mortality: investigating the psychopath. *Cognition* 57, 1–29.

Blair, R. J., and Curran, H. V. (1999). Selective impairment in the recognition of anger induced by diazepam. *Psychopharmacology (Berl.)* 147, 335–338.

Bolger, N., and Zuckerman, A. (1995). A framework for studying personality in the stress process. *J. Pers. Soc. Psychol.* 69, 890–902.

Borkenau, P., and Ostendorf, F. (1993). NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae (S. 5-10, 27-28). Göttingen, Hogrefe.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., and Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage* 25, 312–319.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S. E., and Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 17, 875–887.

Bryden, M. P. (1970). Laterality effects in dichotic listening: relations with handedness and reading ability in children. *Neuropsychologia* 8, 443–450.

Buccino, G., Binkofski, F., and Riggio, L. (2004). The mirror neuron system and action recognition. *Brain Lang.* 89, 370–376.

Calder, A. J., Young, A. W., Rowland, D., Perrett, D. I., Hodges, J. R., and Etcoff, N. L. (1996). Facial emotion recognition after bilateral amygdala damage: differentially severe impairment of fear. *Cogn. Neuropsychol.* 13, 699–745.

Carver, C. S., and White, T. (1994). Behavioural inhibition, behavioural activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Pers. Soc. Psychol. Bull.* 17, 227–232.

Costa, P. T., Jr., and McCrae, R. R. (1992). The Revised NEO Personality Inventory (NEO PI-R) Professional Manual. Odessa, FL, Psychological Assessment Resources.

Craig, E. D. (1968). Physiological arousal as a function of imagined, vicarious, and direct stress experiences. *J. Abnorm. Psychol.* 73, 513–520.

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.* 7, 189–195.

Dapretto, M., and Bookheimer, S. Y. (1999). Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24, 427–432.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Cat. Sel. Doc. Psychol.* 10, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126.

Davis, M. H. (1994). Empathy: A Social Psychological Approach. Madison, WCB Brown and Benchmark.

Davis, M. H. (1996). Empathy: A Social Psychological Approach. Madison, WCB Brown and Benchmark.

Davis, M. H., and Franzoi, S. L. (1991). Stability and change in adolescent self-consciousness and empathy. *J. Res. Pers.* 25, 70–87.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., and Hadjikhani, N. (2004). Fear fosters flight: a mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16701–16706.

de Vignemont, F., and Singer, T. (2006). The empathic brain: how, when and why? *Trends Cogn. Sci. (Regul. Ed.)* 10, 435–441.

de Waal, F. B. M. (1996). Good Natured: The Origins of Right and Wrong in Humans and Other Animals. Cambridge, MA, Harvard University Press.

Decety, J., and Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 71–100.

Decety, J., and Jackson, P. L. (2006). A social neuroscience perspective on empathy. *Curr. Dir. Psychol. Sci.* 15, 54–58.

Decety, J., Jackson, P. L., and Brunet, E. (2007). The cognitive neuropsychology of empathy. In Empathy in Mental Illness and Health, T. F. Farrow and P. W. Woodruff, eds (Cambridge, Cambridge University Press), pp. 239–260.

Decety, J., and Meyer, M. (2008). From emotion resonance to empathic understanding: a social developmental neuroscience account. *Dev. Psychopathol.* 20, 1053–1080.

Decety, J., and Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions. *Biopsychosoc. Med.* 1, 22.

Eippert, F., Veit, R., Weiskopf, N., Erb, M., Birbaumer, N., and Anders, S. (2007). Regulation of emotional responses elicited by threat-related stimuli. *Hum. Brain Mapp.* 28, 409–423.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annu. Rev. Psychol.* 51, 665–697.

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P., Smith, M., O'Boyle, G., and Suh, K. (1994). The relations of emotionality and regulation to dispositional and situational empathy-related responding. *J. Pers. Soc. Psychol.* 66, 776–797.

Eisenberg, N., and Strayer, J. eds (1987). Empathy: A Social Psychological Approach. Dubuque, IA, Brown & Benchmark.

Engels, A. S., Heller, W., Mohanty, A., Herrington, J. D., Banich, M. T., Webb, A. G., and Miller, A. G. (2007). Specificity of regional brain activity in anxiety types during emotion processing. *Psychophysiology* 44, 352–363.

Englis, B. G., Vaughan, K. B., and Lanzetta, J. T. (1982). Conditioning of counter-empathetic emotional responses. *J. Exp. Soc. Psychol.* 18, 375–391.

Fecteau, S., Armony, J. L., Joanette, Y., and Belin, P. (2005). Judgment of emotional nonlinguistic vocalizations: age-related differences. *Appl. Neuropsychol.* 12, 40–48.

Fowles, D. C. (1980). The three arousal model: implications of gray's two-factor learning theory for heart rate, electrodermal activity, and psychopathy. *Psychophysiology* 17, 87–104.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci. (Regul. Ed.)* 8, 396–403.

Gray, J. (1982). The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System. Oxford, Oxford University Press.

Gray, J. A. (1987). The Psychology of Fear and Stress. Cambridge, Cambridge University Press.

Gray, J. A., and McNaughton, N. (2000). The Neuropsychology of Anxiety. London, Oxford University Press.

Grezes, J., Pichon, S., and de Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage* 35, 959–967.

Gu, X., and Han, S. (2007). Attention and reality constraints on the neural processes of empathy for pain. *Neuroimage* 36, 256–267.

Halgren, E., Walter, R. D., Cherlow, D. G., and Crandall, P. H. (1978). Mental phenomena evoked by electrical stimulation of the human hippocampal formation and amygdala. *Brain* 101, 83–117.

Hamann, S., and Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport* 13, 15–19.

Hariri, A. R., Bookheimer, S. Y., and Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport* 11, 43–48.

Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., and Weinberger, D. R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. *Biol. Psychiatry* 53, 494–501.

Hartig, J., and Moosbrugger, H. (2003). Die "ARES-Skalen" zur Erfassung der individuellen BIS- und BAS-Sensitivität: Entwicklung einer Lang- und einer Kurzfassung. *Z. Differ. Diagn. Psychol.* 24, 291–308.

Isenberg, N., Silbersweig, D., Engelien, A., Emmerich, S., Malavade, K., Beattie, B., Leon, A. C., and Stern, E. (1999). Linguistic threat activates the human amygdala. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10456–10459.

Izard, C. E., Libero, D. Z., Putnam, P., and Haynes, O. M. (1993). Stability of emotion experiences and their relations to traits of personality. *J. Pers. Soc. Psychol.* 64, 847–860.

Jackson, P. L., Brunet, E., Meltzoff, A. N., and Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia* 44, 752–761.

Jackson, P. L., Meltzoff, A. N., and Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage* 24, 771–779.

Jackson, P. L., Meltzoff, A. N., and Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31, 429–439.

Jäncke, L., Cheetham, M., and Baumgartner, T. (2009). Virtual reality and the role of the prefrontal cortex in adults and children. *Front. Hum. Neurosci.* 3, 52–59. doi: 10.3389/neuro.01.006.2009.

Lamm, C., Batson, C. D., and Decety, J. (2007a). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *J. Cogn. Neurosci.* 19, 42–58.

Lamm, C., Nusbaum, H. C., Meltzoff, A. N., and Decety, J. (2007b). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS ONE* 2, e1292. doi: 10.1371/journal.pone.0001292.

Lanzetta, J. T., and Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *J. Pers. Soc. Psychol.* 56, 543–554.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., and Way, B. M. (2007). Putting feelings into words: affect labeling disrupts amygdala activity in response to affective stimuli. *Psychol. Sci.* 18, 421–428.

McCrae, R. R., and Costa, P. T., Jr. (1991). Adding liebe und arbeit: the full five-factor model and well-being. *Pers. Soc. Psychol. Bull.* 17, 227–232.

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Psychol.* 67, 371–378.

Milgram, S. (1974). Obedience to Authority. New York, Harper and Row.

Morrison, I., and Downing, P. E. (2007). Organization of felt and seen pain responses in anterior cingulate cortex. *Neuroimage* 37, 642–651.

Morrison, I., Lloyd, D., di Pellegrino, G., and Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? *Cogn. Affect. Behav. Neurosci.* 4, 270–278.

Morrison, I., Peelen, M. V., and Downing, P. E. (2007). The sight of others' pain modulates motor processing in human cingulate cortex. *Cereb. Cortex* 17, 2214–2222.

Nomura, M., Ohira, H., Haneda, K., Iidaka, T., Sadato, N., Okada, T., and Yonekura, Y. (2004). Functional association of the amygdala and ventral prefrontal cortex during cognitive evaluation of facial expressions primed by masked angry faces: an event-related fMRI study. *Neuroimage* 21, 352–363.

Ochsner, K. N., Bunge, S. A., Gross, J. J., and Gabrieli, J. D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *J. Cogn. Neurosci.* 14, 1215–1229.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., and Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage* 23, 483–499.

Oliveri, M., Babiloni, C., Filippi, M. M., Caltagirone, C., Babiloni, F., Cicinelli, P., Traversa, R., Palmieri, M., and Rossini, P. M. (2003). Influence of the supplementary motor area on primary motor cortex excitability during movements triggered by neutral or emotionally unpleasant visual cues. *Exp. Brain Res.* 149, 214–221.

Paulus, C. (1992). Empathie, Kompetenz und Altruismus. Available at: www.uni-saarland.de/ak5/ezw/ abteil/motiv/paper/empathie.htm.

Petersen, S. E., Fox, P. T., Snyder, A. Z., and Raichle, M. E. (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science* 249, 1041–1044.

Petrovic, P., Carlsson, K., Petersson, K. M., Hansson, P., and Ingvar, M. (2004). Context-dependent deactivation of the amygdala during pain. *J. Cogn. Neurosci.* 16, 1289–1301.

Petrovic, P., and Ingvar, M. (2002). Imaging cognitive modulation of pain processing. *Pain* 95, 1–5.

Petrovic, P., Ingvar, M., Stone-Elander, S., Petersson, K. M., and Hansson, P. (1999). A PET activation study of dynamic mechanical allodynia in patients with mononeuropathy. *Pain* 83, 459–470.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., Williams, V. S. C. R., Bullmore, E. T., Brammer, M., and Gray, J. A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc. Biol. Sci.* 265, 1809–1817.

Pichon, S., de Gelder, B., and Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Soc. Neurosci.* 3, 199–212.

Pickering, A. D., and Gray, J. A. (2001). Dopamine, appetitive reinforcement and the neuropsychology of human learning: an individual differences approach. In Advances in Individual Differences Research, A. Eliasz and A. Angleitner, eds (Lengerich, PABST Science Publishers), pp. 113–149.

Preston, S. D., and de Waal, F. B. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20; discussion 20–71.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. (1999). SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* 42, 952–962.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670.

Saarela, M. V., Hlushchuk, Y., Williams, A. C., Schurmann, M., Kalso, E., and Hari, R. (2007). The compassionate brain: humans detect intensity of pain from another's face. *Cereb. Cortex* 17, 230–237.

Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 332–339.

Sander, K., and Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res. Cogn. Brain Res.* 12, 181–198.

Schaefer, S. M., Jackson, D. C., Davidson, R. J., Aguirre, G. K., Kimberg, D. Y., and Thompson-Schill, S. L. (2002). Modulation of amygdalar activity by the conscious regulation of negative emotion. *J. Cogn. Neurosci.* 14, 913–921.

Schienle, A., Schafer, A., Hermann, A., Walter, B., Stark, R., and Vaitl, D. (2006). fMRI responses to pictures of mutilation and contamination. *Neurosci. Lett.* 393, 174–178.

Schulz, G. M., Varga, M., Jeffires, K., Ludlow, C. L., and Braun, A. R. (2005). Functional neuroanatomy of human vocalization: an H215O PET study. *Cereb. Cortex* 15, 1835–1847.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., and Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385, 254–257.

Simon, D., Craig, K. D., Miltner, W. H., and Rainville, P. (2006). Brain responses to dynamic facial expressions of pain. *Pain* 126, 309–318.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439, 466–469.

Sinha, R. (2001). How does stress increase risk of drug abuse and relapse?

Psychopharmacology (Berl.) 158, 343–359.

Slater, M. (2002). Presence and the sixth sense. *Presence* 11, 435–439.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc.* Available at: http://www.cs.ucl.ac.uk/staff/m.slater/Papers/rss-prepublication.pdf/

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., and Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE* 1, e39. doi: 10.1371/journal.pone.0000039.

Steen, F. F., and Owens, S. (2001). Evolution's pedagogy: an adaptationist model of pretense and entertainment. *J. Cogn. Cult.* 1, 289–321.

Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R.,

Stirling, J., McKie, S., Deakin, J. F. W., and Elliott, R. (2006). Neural correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage* 29, 90–98.

Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., Kosslyn, S. M., Rose, R. M., and Cohen, J. D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science* 303, 1162–1167.

Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J., Rosen, S., and Scott, S. K. (2006). Positive emotions preferentially engage an auditory-motor "mirror" system. *J. Neurosci.* 26, 13067–13075.

Watson, D., and Clark, L. A. (1992). On traits and temperament: general and specific factors of emotional experi-

ence and their relation to the five-factor model. *J. Pers.* 60, 441–476.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., and Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage* 24, 1233–1241.

Williams, A. C. (2002). Facial expression of pain: an evolutionary account. *Behav. Brain Sci.* 25, 439–455; discussion 455–488.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., and Jäncke, P. (2007). A process model of the formation of spatial presence experiences. *Media Psychol.* 9, 493–525.