

Familiarity promotes the blurring of self and other in the neural representation of threat

Lane Beckes,¹ James A. Coan,¹ and Karen Hasselmo²

¹Department of Psychology, University of Virginia, Charlottesville, VA, USA and ²Department of Psychology, University of Arizona, Tucson, AZ, USA

Neurobiological investigations of empathy often support an embodied simulation account. Using functional magnetic resonance imaging (fMRI), we monitored statistical associations between brain activations indicating self-focused threat to those indicating threats to a familiar friend or an unfamiliar stranger. Results in regions such as the anterior insula, putamen and supramarginal gyrus indicate that self-focused threat activations are robustly correlated with friend-focused threat activations but not stranger-focused threat activations. These results suggest that one of the defining features of human social bonding may be increasing levels of overlap between neural representations of self and other. This article presents a novel and important methodological approach to fMRI empathy studies, which informs how differences in brain activation can be detected in such studies and how covariate approaches can provide novel and important information regarding the brain and empathy.

Keywords: empathy; emotion; social cognition; interpersonal relationships; prosocial behavior; familiarity

INTRODUCTION

Empathy may be critical for understanding others and motivating altruism (Batson and Shaw, 1991). de Vignement and Singer (2006, p. 435) define empathy as having occurred when ‘... (i) one is in an affective state; (ii) this state is isomorphic to another person’s affective state; (iii) this state is elicited by the observation or imagination of another person’s affective state; (iv) one knows that the other person is the source of one’s own affective state’. Studies suggest a consistent set of neural regions similarly responsive to both self- and other-directed cues of threat and pain (Decety, 2011). These regions include the anterior insula (AI), prefrontal cortex (PFC), orbitofrontal cortex (OFC), various subcortical affective regions and the anterior cingulate cortex (ACC). Moreover, activations in these regions are frequently characterized as ‘overlapping’ across self and others, or as providing evidence of ‘self-other’ overlap (Decety and Sommerville, 2003; de Vignement and Singer, 2006;). Despite their consistency, these effects are moderated by other variables (cf. Hein and Singer, 2008). For example, physicians have generally weaker responses to vicarious pain (Cheng *et al.*, 2007), and men respond less to the pain of unfair individuals (Singer *et al.*, 2006). Furthermore, similar pain in targets produces stronger responses than dissimilar pain (Lamm *et al.*, 2010). In this article, we compare degrees of self–other overlap in threat-responsive brain regions as a function of familiarity and add to traditional data analytic approaches to these questions an alternative strategy designed to highlight individual differences in self–other overlap and its moderation by familiarity.

Simulation theory and conjunction analysis

Neuroimaging studies suggest that empathy involves simulating the experience of others using neural circuits dedicated to perceiving the

state of one’s own body (e.g. Carr *et al.*, 2003; Singer *et al.*, 2004; Lamm *et al.*, 2007, 2010; Ochsner *et al.*, 2008; Keysers and Gazzola, 2007), a pattern broadly predicted by simulation theory (Gallese and Goldman, 1998). For example, Singer *et al.* (2004) discovered that neural circuits supporting affective responses to one’s own pain are similarly active during another’s pain. Moreover, activation intensity in these circuits during other-directed pain varies in part as a function of self-reported empathy.

Researchers typically use conjunction analysis to identify shared neural networks. Conjunction analysis determines brain regions that are active in two or more conditions. In the typical empathy study, these would be areas that are active when researchers apply an aversive stimulus both to the participant and to someone else—the latter of which putatively indicates an empathic response. Activations during empathic responding that mirror self-focused responses are interpreted as reflecting the use of one’s own experience to simulate the other’s psychological state, a process that Singer *et al.* (2004) have suggested implies a ‘breach of individual separateness’. Indeed, conjunction analysis is frequently interpreted as suggesting that self-related brain activation is correlated with other-related brain activity when people engage in empathy. For example, Decety (2011, p. 104) argues that ‘... similar neural networks mediate the simulation of pain for self and other. Such a perception–action coupling mechanism offers an interesting foundation for intersubjectivity because it provides a functional bridge between first-person information and third-person information, grounded on self–other equivalence...’. Although these interpretations imply *correlation* between self and other, conjunction analysis does not.¹ There may be no intra-individual correlation in a given region between the representation of self and of other even when overall average group activity in both conditions is similar. For

Received 16 September 2011; Accepted 16 April 2012

Advance Access publication 3 May 2012

The authors acknowledge the support of Alexander Tatum, Amanda LeTard, Casey Brown, Matthew Allen, Thomas Hale-Kupiec, Caroline Trower, Priyanka Banjeree, Joe Allen, Jon Haidt, James Morris and Jeff Simpson. They also thank Joe Allen for access to the KLIFF sample.

This project was supported by a grant issued by the National Institute of Mental Health. The project described was supported by Award Number R01MH080725 to J.A.C. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Correspondence should be addressed to James A. Coan, Department of Psychology, University of Virginia, 102 Gilmer Hall, Charlottesville, VA 22904, USA. E-mail: jcoan@virginia.edu

¹There may be little or no intra-individual correlation in a given region between the representation of self and other even when overall average activity in both conditions is similar across many individuals. Conversely, there may be a high self–other correlation in a given region even when overall average activity in both conditions is quite different when averaged across individuals. This is because the general order and relative magnitude of each person’s data points are independent of the overall group mean. For example, a vector composed of the numbers 2 4 3 5 1 6 is perfectly correlated with a vector composed of the numbers 12 14 13 15 11 16, but the means of each vector are dramatically different. This analytical difference may have important implications for the construct that our measures are tapping. For example, conjunction analysis likely tells us quite accurately which regions are normatively involved in empathic processing of vicariously experienced negative events, whereas correlation in this case likely reflects the tendency to treat the other person as if they were the self or just like the self, a pattern that may suggest an altogether different psychological construct.

example, people may use mental simulations to predict the behavior or experiences of others in ways that do not suggest a ‘breach of individual separateness’. Conversely, there may be a high self–other correlation in a given region even when overall average activity in both conditions is quite different at the group level.

A data analytic approach that combines the group and individual levels of analysis would provide a richer understanding of the neural processes underpinning empathy—broadly speaking—and simulation more specifically. Although conjunction analysis is restricted to the group level of analysis, the statistical approach we take allows for detailed analyses of effects at both the group and individual levels. Applying both approaches, we first replicated previously reported conjunction findings suggesting that similar or identical threat-related activations obtain at the group level regardless of whether the threat is directed at the self, a friend or a stranger (as in Figure 1). In contrast, a mixed model approach that included both group and individual levels of analysis suggested statistically significant differences in the degree of threat-related self–other overlap, depending on who the threat was directed at. Specifically, we observed that self–friend correlations were high throughout the threat-responsive brain, whereas self–stranger correlations were comparatively quite low (as in Figure 2B). This research clarifies the ways in which people represent the thoughts and feelings of others within the brain, offers insights into analytic approaches for investigating these processes and provides a more nuanced view concerning how the human brain understands the minds of others.

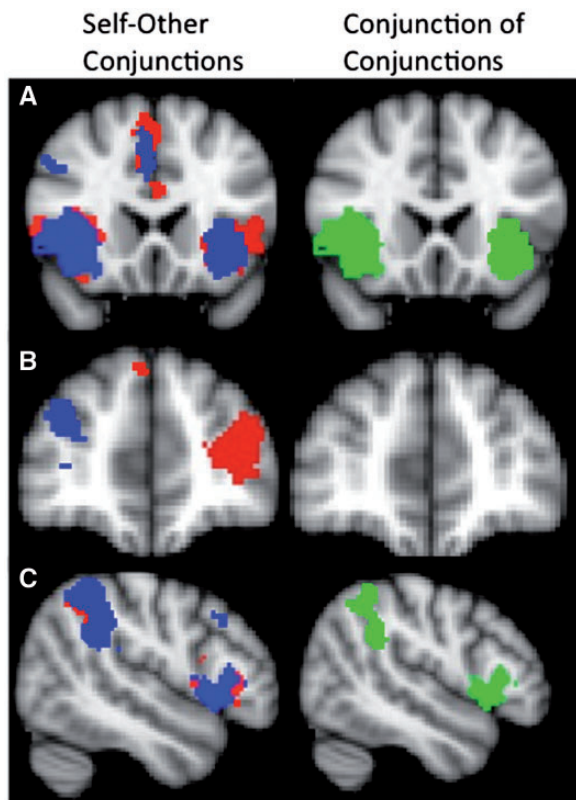


Fig. 1 Significant conjunctions between the threat-to-self and threat-to-other conditions on the left, with self–friend overlap indicated in red and self–stranger overlap indicated in blue. On the right, the ‘conjunction of conjunctions’ depicts areas (green) where self–friend and self–stranger conjunctions are themselves conjuncted. (A) Conjunctions in portions of the ACC and AI. (B) The most prominent differences between the two conjunction analyses, showing significant self–friend overlap in the left PFC and significant self–stranger overlap in the right PFC. (C) Conjunctions in portions of the right SMG.

METHODS

Participants

Twenty-five participants brought an opposite-gender friend to the scan. Three pairs were dropped due to (i) a technical issue, (ii) bringing in a sibling and (iii) being outliers according to mahalabonis distances (results are unchanged). The mean age of the 22 final scanned participants (11 women) was 23.59 years (s.d. = 0.959). Fourteen participants identified as white and eight as African-American. Participants were recruited from the KLIFF sample, which the Allen Laboratory (Allen *et al.*, 2007; McElhaney *et al.*, 2008) has been assessing for over a decade. KLIFF participants were scanned, and friends provided hand holding. Informed consent was obtained from both members of each pair, and participants were paid \$160 each for participation.

Measures

Inclusion of the Other in the Self Scale

The inclusion of the Other in the Self (IOS) scale is a Venn diagram depicting seven overlapping pairs of circles. On one end of the continuum, the circles are completely separate, and on the other end, the circles are virtually entirely overlapping. The measure is designed to tap the extent to which the participant is interconnected with the other person of interest. Aron *et al.* (1992) found that the scale had good psychometric properties including test–retest reliability, convergent validity by correlations with measures of relationship closeness, and predictive validity by subsequent measures of relationship maintenance.

Procedure

Participants completed questionnaires and practiced the functional magnetic resonance imaging (fMRI) procedure. Two Ag–AgCl shock electrodes were applied to the participant’s ankle (left or right counter-balanced across participants) and their friend’s ankle. The KLIFF member was then taken into the scanning chamber where high-resolution anatomical scans followed the practice session.

Participants viewed stimuli projected onto a screen behind the magnet’s bore through a mirror and responded to stimuli by button box. The study consisted of five experimental blocks, during which the participant viewed 10 threat cues with no shock, 2 with shock and 12 safety cues in variable order. The first two scanning blocks made up the ‘Threat to Other’ portion of the study, where mild electric shock was delivered to the person the participant was holding hands with. ‘Threat to Self’ composed the final three scanning blocks, where mild electric shock was delivered to the participant while they were either holding hands with the friend, a stranger or no one (Coan *et al.*, 2006). All subjects experienced all five blocks. Trials were varied within subject, and block order was counterbalanced between subjects. The stranger was an anonymous member of the opposite gender. Participants’ right hands were used for hand holding, whereas their left hand held the button box. Threat cues consisted of a red ‘X’ on a black background and indicated a 17% chance of electric shock. Safety cues, a blue ‘O’ on a black background, indicated no chance of shock. Shocks were generated by an isolated physiological stimulator (Coulbourn Instruments, Allentown, PA, USA) and lasted for 20 ms at 4 mA.

Trials were composed of a 1 s safety/threat cue, followed by 4–10 s of anticipation indicated by a fixation cross. Subsequently, a small dot appeared, during which shocks were delivered if a shock trial. The inter-trial interval was 4 to 10 s. Participants rated their subjective feelings of unpleasantness (valence) and agitation (arousal) on the Self-Assessment Manikin (SAM) scales (Bradley and Lang, 1994), a 9-point pictorial scale, after each block.

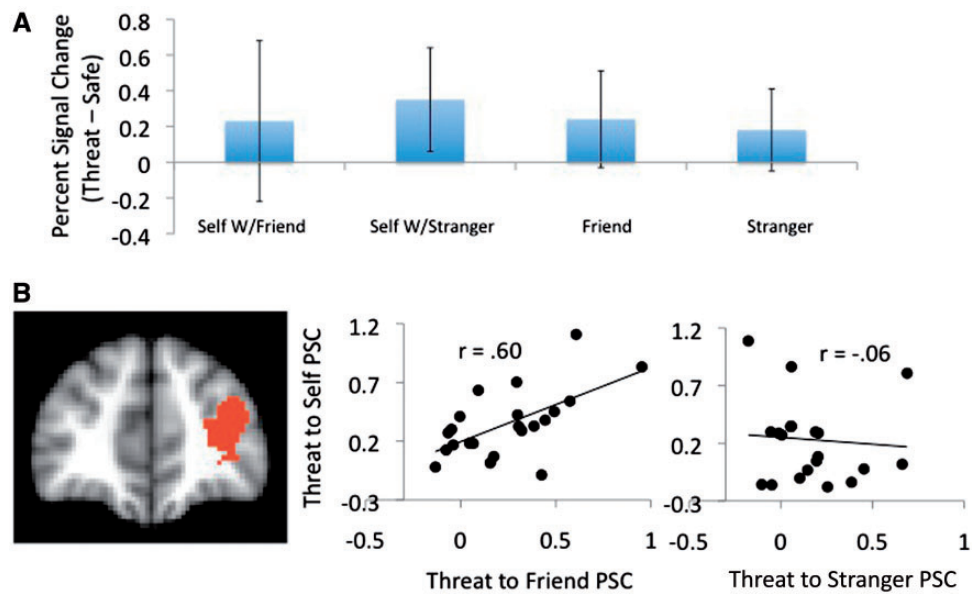


Fig. 2 (A) Bar graphs showing group level means and standard deviations of PSC by condition in the left LPFC ($-34.4, 16.3, -2.46$). (B) Scatter plots showing the correlation between threat-to-self and threat-to-other PSC in the same region. Note the lack of apparent differences in the group means (A), despite dramatic differences in the correlations between self-threat PSC and other-threat PSC as a function of familiarity (B).

Image acquisition and data analysis

Functional images were acquired using a Siemens 3.0 Tesla MAGNETOM Trio high-speed magnetic imaging device at University of Virginia's Fontaine Research Park, with a circularly polarized the transmit/receive head coil with integrated mirror. A total of 216 functional T2*-weighted echo planar images (EPIS) sensitive to blood-oxygen-level-dependent contrast were collected per block, in volumes of 28 3.5-mm transversal echo-planar slices (1-mm slice gap) covering the whole brain (1-mm slice gap, repetition time (TR)=2000 ms, echo time (TE)=40 ms, flip angle = 90° , field of view (FOV) = 192 mm, matrix = 64×64 , voxel size = $3 \times 3 \times 3.5$ mm). Before collection of functional images, 176 high-resolution T1-magnetization-prepared rapid-acquisition gradient echo images were acquired to determine the localization of function (1-mm slices, TR = 1900 ms, TE = 2.53 ms, flip angle = 9° , FOV = 250 mm, voxel size = $1 \times 1 \times 1$ mm).

Data were preprocessed and analyzed using FMRIB's Software Library (FSL) software (Version 5.98; www.fmrib.ox.ac.uk/fsl, Worsley, 1994). Motion correction involved FMRIB's Linear Image Registration Tool, an intra-modal correction algorithm tool (MCFLIRT; Jenkinson *et al.*, 2002), with slice scan time correction and a high-pass filtering cutoff point of 100 s, removing irrelevant signals. We used BET (Smith, 2002) brain extraction, eliminating non-brain material voxels in the fMRI data, and a 5-mm full width at half minimum Gaussian kernel for smoothing. Images were registered to the Montreal Neurological Institute (MNI) space by FLIRT (Jenkinson *et al.*, 2002). Trials in which participants received shocks were excluded due to movement artifacts and our primary interest in anticipatory threat.

Conjunction analyses

We conducted conjunction analyses using the `easythresh_conj` command within FSL (Nichols, 2007). This command detects areas of significant activation in two or more conditions by testing the 'conjunction null hypothesis'. This statistical procedure tests whether activation is significant in both of two conditions. All conjunctions were of images

created with a non-masked whole-brain analysis of the threat minus safe contrast, within conditions using the standard z threshold of 2.3 and a P threshold of 0.05. Conjoined analyses included the self- and friend-directed threat conditions, and the self- and stranger-directed threat conditions. Finally, we conducted a 'conjunction of the conjunctions'. This third analysis mapped the spatial extent of the similarity between self-friend and self-stranger threat-safe conjunctions.

Functional regions of interest

We first characterized the neural threat response using a contrast of threat minus safe derived from the alone, threat-to-self condition to localize threat-related activity (cf. Coan *et al.*, 2006; Poldrack, 2007). First-level analysis of the functional data began with the determination of functional regions of interest (ROIs) using FEAT and time-series statistical analysis by FILM (Worsley, 2001). Higher level analysis was performed by FLAME (FMRIB's Local Analysis of Mixed Effects) state 1. Multisubject ROIs were identified by cluster-wise tests using the FSL standard z threshold of 2.3 and cluster P threshold of 0.05. These ROIs were then used to create structural masks using FSLView's Harvard-Oxford Cortical and Subcortical atlases, revealing activations frequently associated with the neural response to threat (cf. Coan *et al.*, 2006; Table 1). Applying these masks to all other conditions (using the threat minus safe contrast), parameter estimates were extracted from each ROI, for each subject, using FEATQuery. These estimates were converted to mean percent signal change (PSC) values across all voxels, providing mean ROI PSC estimates for each subject and each experimental condition. Estimates were then input into the PASW statistical package, version 18, for linear mixed models (LMMs) and correlational analyses reported earlier.

RESULTS

Conjunction analysis

Conjunction analyses revealed highly similar activations throughout the brain, regardless of whether threats were directed at the self, the friend or the stranger (Figure 1). Few differences suggested a familiarity effect. Conjunctions specific to the self and stranger conditions were

Table 1 Functional ROI, with coordinates, maxima and cluster sizes

	Centroid coordinates			z max	Voxel size
	x	y	z		
Frontal regions					
Left OFC	−34	24	−8	4.26	669
Right OFC	38	24	−6	4.76	839
Left LPFC	−34	46	14	3.62	542
Right LPFC	36	44	18	4.07	1550
Left SupFG	−2	22	50	3.27	157
Right SupFG	6	26	48	3.96	643
Left SMC	−6	2	56	3.65	352
Right SMC	6	4	56	3.24	229
Left IFG	−50	10	0	3.37	287
Right IFG	52	18	4	4.06	491
Other cortical regions					
Right SMG	52	−44	38	4.67	1603
Left insula	−34	16	−2	4.26	799
Right insula	36	20	−2	4.76	674
PCC	2	−28	24	3.60	376
Dorsal ACC	2	20	32	2.94	1208
Subcortical regions					
Left putamen	−26	8	0	3.51	351
Right putamen	28	18	0	3.09	266
Left thalamus	−6	−10	−4	3.01	28
Right thalamus	8	−4	6	3.74	272
Right caudate	14	6	10	3.92	305

observed in the right lateral prefrontal cortex (LPFC), whereas conjunctions specific to the self and friend conditions were observed in the left LPFC and left putamen. Moreover, the spatial extent of the self–friend conjunction in the ACC region appeared to be greater than that of the self–stranger conjunction (1237 voxels *vs* 967 voxels). Outside of these differences, all remaining self–other conjunctions were virtually identical. A ‘conjunction of conjunctions’, where self–friend conjunctions are themselves conjoined with self–stranger conjunctions, revealed self–other overlap in the in ACC, AI and right supramarginal gyrus (SMG) (Figure 1).

Importantly, the apparent differences between friend and stranger conjunctions may not be statistically meaningful. As reviewed in our Supplementary Materials, we observed no significant differences between the threat-to-friend and threat-to-stranger conditions using a whole-brain higher-level analysis to test for familiarity effects (Supplementary Figure 1). Although the conjunction analysis can identify where two conditions share significant activations relative to baseline, it cannot determine whether activation in one non-baseline condition is significantly different from another. This renders apparent differences in observed conjunctions between the threat-to-friend and threat-to-stranger conditions difficult to interpret.

Covariate models of self–other overlap using functional ROIs

Threat-to-other PSC was predicted using an LMM including one categorical familiarity factor (stranger *vs* friend) and one continuous predictor: PSC attributable to the threat directed at the self (threat-to-self). This predictor was a repeated covariate with a data point for each participant corresponding to each of the two familiarity conditions (stranger *vs* friend). Because of the way this analysis was constructed, threat-to-self and threat-to-other were matched for hand holding. Specifically, the score from the threat-to-self friend handholding condition was matched with the score from the threat-to-other friend condition, and the score from

the threat-to-self stranger handholding condition was matched with the score from the threat-to-other stranger condition. Thus, one would interpret an interaction between threat-to-self and familiarity as indicating that the correlation between threat-to-self (friend hand holding) and threat-to-friend is significantly different from the correlation between threat-to-self (stranger hand holding) and threat-to-stranger. The dependent variable in each LMM was PSC in the threat-to-other conditions. Results are listed in Table 2. Zero-order correlations representing self–friend overlap and self–stranger overlap are also included to aid interpretation.

Main effects of familiarity

No main effects of familiarity were found—evidence that at a group-level participants were simulating threats to friends and strangers in approximately equal amounts. To be certain that no differences between threat-to-friend and threat-to-stranger existed an additional whole-brain analysis (with non-masked data) comparing threat-to-friend (threat minus safe) and threat-to-stranger (threat minus safe) indicated that there were no significant areas of difference between the two conditions (see the Supplementary Materials for more information and images). This overall pattern of activation on behalf of others replicates early work (e.g. Singer *et al.*, 2004) but does not speak to the question of whether familiarity moderates the way in which people process threats to others relative to threats to self.

Main effects of threat-to-self

Main effects of threat-to-self PSC on threat-to-other PSC would indicate that threat-to-self activations were correlated with threat-to-other activations regardless of the other’s level of familiarity. Such main effects were observed in OFC, LPFC, right thalamus and left superior frontal gyrus (SFG). In nearly all these regions, with the exception of the right thalamus and right OFC, these main effects seem to be interpretable only in light of the interaction between self-threat PSC and familiarity, discussed later. The pattern indicates large correlations between self- and friend-directed threat, with self- and stranger-directed threat correlations small to negligible. A notable exception to this is in the right OFC, which is commonly involved in representations of stimulus value (Berkman and Lieberman, 2009) and likely indicates that participants were encoding threat cues as generally ‘negative’ for both friends and strangers.

Interactions between familiarity and threat-to-self PSC

Familiarity by threat-to-self PSC interaction effects were observed in the putamen, LPFC, left OFC, left SFG and right SMG. Subsequent decompositions indicated significant self–friend correlations in contrast to little or no self–stranger correlation in neural threat representation (Table 2, Figures 2 and 3). Several of these effects survived Bonferroni correction (which is generally considered overly conservative in correlated data, cf. Williams *et al.*, 1999; Conneely and Boehnke, 2007), including the LPFC and right putamen. Additionally, whole brain corrected analyses on each threat-to-other data set using threat-to-self as a voxel-dependent predictor revealed extensive areas of correlation between self and friend and little to no correlation between self and stranger (Supplementary Materials).

Arousal and valence reports

Subjective reports of affective arousal and valence replicated the pattern found in the brain correlation data. An analysis of covariance (ANCOVA) indicated that participants’ valence ratings after self-threat

Table 2 Functional ROI, *F* tests, parameter estimates and self–other time-series correlations

Region	Self–friend correlation	Self–stranger correlation	Main effect familiarity			Main effect threat-to-self			Familiarity by threat-to-self interaction		
			<i>t</i>	<i>P</i>	P.E.	<i>t</i>	<i>P</i>	P.E.	<i>t</i>	<i>P</i>	P.E.
Frontal regions											
Left OFC	0.60 ^a	0.16	0.15	0.90	0.005	2.4	0.02	0.126	2.7	0.01	0.027
Right OFC	0.59 ^a	0.43 ^b	1.0	0.34	0.031	3.2	0.003	0.147	1.9	0.07	0.071
Left LPFC	0.60 ^a	−0.06	0.3	0.6	0.014	2.5	0.02	0.096	3.4	0.002	0.110
Right LPFC	0.66 ^c	−0.22	1.1	0.3	0.027	2.3	0.03	8.6	4.1	<0.001	0.131
Left SupFG	0.54 ^d	0.12	1.0	.32	0.039	2.2	0.03	0.111	2.2	0.04	0.098
Right SupFG	0.18	0.06	1.4	0.18	0.041	0.1	0.94	0.004	0.8	0.45	0.030
Left IFG	0.31	0.10	0.8	0.42	0.033	0.5	0.65	0.029	0.7	0.5	0.039
Right IFG	0.41 ^b	−0.10	0.7	0.48	0.028	0.0	0.96	0.002	1.3	0.20	0.060
Left SMC	0.26	0.18	0.4	0.73	0.014	1.4	0.17	0.066	0.7	0.50	0.031
Right SMC	0.23	0.09	−0.6	0.57	−0.022	0.7	0.50	0.039	0.9	0.38	0.042
Other cortical regions											
Right SMG	0.63 ^a	−0.29	0.8	0.46	0.020	0.7	0.49	0.031	2.6	0.02	0.089
Left anterior insula	0.40+	0.07	−0.3	0.80	−0.011	1.4	0.18	0.075	1.2	0.24	0.057
Right anterior insula	0.40+	−0.04	0.26	0.80	0.008	0.8	0.44	0.033	2.0	0.06	0.068
PCC	0.02	−0.29	3.0	0.10	−0.055	0.9	0.36	−0.038	0.43	0.52	0.026
Dorsal ACC	0.27	0.16	−0.2	0.81	−0.011	1.2	0.23	0.065	0.51	0.61	0.025
Subcortical regions											
Left putamen	0.51 ^b	−0.16	1.3	0.22	0.030	1.1	0.31	0.034	3.1	0.005	0.088
Right putamen	0.53 ^d	−0.23	0.7	0.5	0.014	1.3	0.19	0.037	3.8	0.001	0.097
Left thalamus	0.28	0.12	−0.1	0.90	−0.004	1.5	0.15	0.054	0.5	0.60	0.020
Right thalamus	−0.21	−0.19	0.2	0.85	0.008	−2.2	0.04	−0.121	−1.2	0.23	−0.066
Right caudate	0.25	−0.34	1.3	0.22	0.038	−0.6	0.57	−0.022	1.6	0.12	0.058

The bottom of the table shows the location in MNI coordinates of the peak voxel within each cluster. + correlation is marginally significant at the $P < 0.10$ level. P.E., parameter estimate.

^aCorrelation is significant at the $P < 0.005$ level.

^bCorrelation is significant at the $P < 0.05$ level.

^cCorrelation is significant at the $P < 0.001$ level.

^dCorrelation is significant at the $P < 0.01$ level.

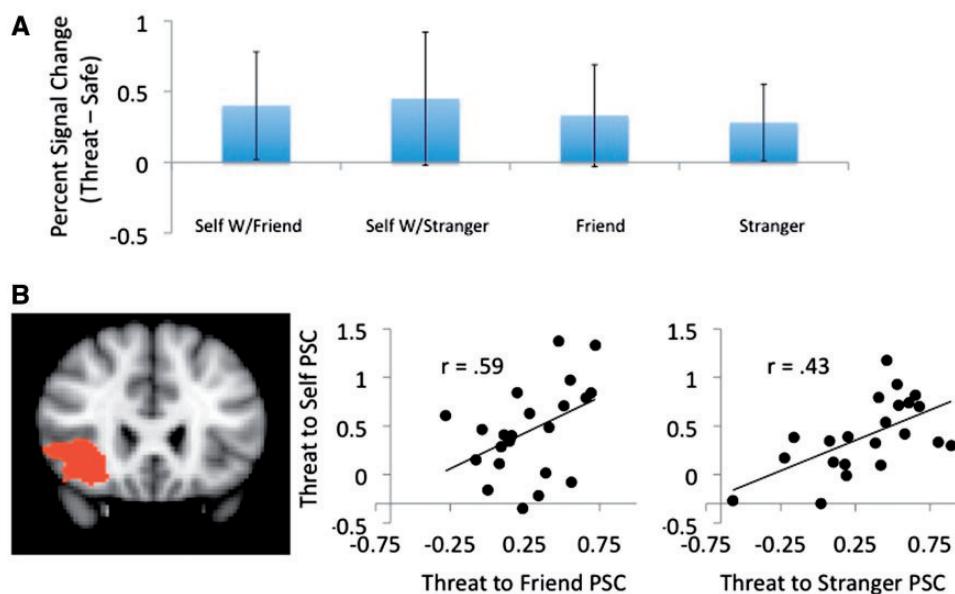


Fig. 3 (A) Bar graphs showing group level means and standard deviations of PSC by condition in the right OFC (37.4, 24.9, −5.6). (B) Scatter plots showing the correlation between threat-to-self and threat-to-other PSC in the same region. Note that here both the group means and individual differences suggest substantial similarity in how the right OFC is processing self-, friend- and stranger-directed threats.

interacted with familiarity to predict their other-threat valence ratings, $F(1, 17) = 4.71$, $P = 0.04$. Specifically, self-threat valence ratings correlated positively with friend-threat ratings ($r = 0.30$) but not with stranger-threat ratings ($r = -0.08$). Self-reported arousal followed a similar pattern, however, self-threat ratings of arousal correlated with both friend-threat ($r = 0.72$) and stranger-threat ratings (0.48), indicated by a main effect of self-threat, $F(1, 17) = 17.04$, $P = 0.001$. This

suggests more similar subjective responses to threat-to-self and threat-to-friend than threat-to-stranger.

Inclusion of IOS scale

To test our interpretation of the main analyses, we also conducted ANCOVAs on both the threat-to-friend and threat-to-stranger

Table 3 Functional ROI, centroid coordinates, z score of the peak voxel and size of the cluster in voxels with IOS scale as a covariate predictor and threat-to-friend (threat minus safe contrast) as the dependent variable

	Centroid coordinates			z max	Size vox.
	x	y	z		
Clusters					
Middle Frontal Gyrus (MFG)/ Inferior Frontal Gyrus (IFG)	46	22	24	3.80	623
OFC	50	22	−6	3.90	371
	Peak coordinates				
	x	y	z		
IFG	56	32	16	–	–
IFG	54	20	−6	–	–

imaging data, with the IOS scale (Aron *et al.*, 1992) as a covariate using the FSL standard z threshold of 2.3 and cluster *P* threshold of 0.05. As can be seen in Table 3 and Figure 4, the IOS scale was robustly correlated with threat activation during the threat-to-friend condition in the right middle and inferior frontal gyrus, right OFC and right anterior insula. In contrast, no correlations between IOS scale and threat-to-stranger were detected. This pattern strongly supports the conclusion that familiarity promotes levels of self–other overlap in neural processing.

DISCUSSION

Replications

Our conjunction results replicated numerous previous findings in the neuroimaging literature on empathy (e.g. Singer *et al.*, 2004; Lamm *et al.*, 2007, 2010; Ochsner *et al.*, 2008). Threats directed at the self, friends and strangers all activated regions previously identified as involved in both threat responding and empathy. Singer and Lamm (2009) have suggested that the empathy network in particular includes structures involved in top–down support of situational appraisal, attention and executive functions, and bottom–up mechanisms supporting affective responses. Our results are consonant with their characterization, suggesting the involvement of structures supporting ‘top down’ processes such as the LPFC, OFC and SFG and those involved in more ‘bottom up’ affective processes such as the putamen.

Covariate models vs conjunction analysis

The use of covariate models to identify the effect of familiarity on empathy produced a set of results that were, however, strikingly different than those derived from our conjunction analysis. Thus, conclusions about the nature of self–other overlap in neural representations of threat may depend non-trivially on the data analytic strategy used. Although conjunction analyses did identify most of the regions where threat-related activations occurred similarly across self, friend and stranger conditions, they nevertheless masked the effects of familiarity revealed by our covariate models. Indeed, our covariate models revealed numerous regions in which threat-to-self activation was moderately to highly correlated with threat-to-friend activation but only weakly correlated with threat-to-stranger activation.

Past studies have indicated that shared networks are involved in the processing of self and other but not in terms of the degree of similarity at the level of individual differences. Although factors such as the perceived fairness of a target or the oddness of a target’s pain modulate the mean activity within these regions (Singer *et al.*, 2006; Lamm *et al.*, 2010), familiarity with the target seems to modulate the correlation between self and other in ways that are not detectable at the level of

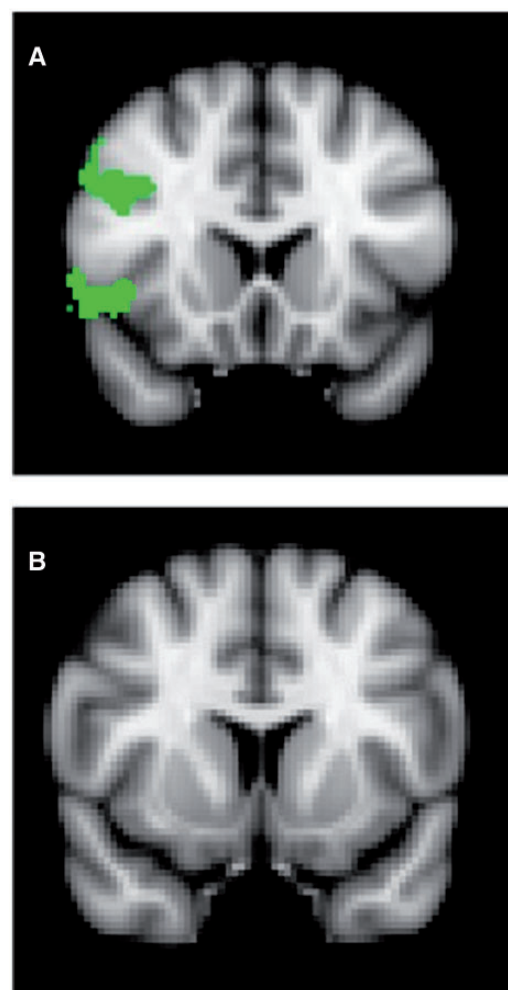


Fig. 4 The significant clusters of activation in (A) a covariance analysis with IOS scale as a covariate predictor and threat-to-friend (threat minus safe contrast) as the dependent variable, and the lack of significant activation in (B) a covariance analysis with IOS scale as a covariate predictor and threat-to-stranger (threat minus safe contrast) as a dependent variable.

group averages. Particularly noteworthy are instances in which conjunction analysis would lead to different inferences than the covariate models. For example, in the right LPFC and right SMG, conjunction revealed significant overlap in activation for the self and stranger conditions but not the self and friend conditions. Our covariate model, however, suggests something like the opposite—that in both of these areas, threat-to-self activation was significantly and positively correlated with threat-to-friend activation but not threat-to-stranger activation. These seemingly contradictory results are understandable only in light of the data analytic differences between conjunction analysis and our covariate model. Conjunction analysis is sensitive only to group-level activation intensity, and our conjunction analyses indeed suggest significant coactivation across a variety of regions when comparing self-directed threats to threats directed at both friends and strangers. However, our covariate models were capable of estimating activation intensity at both the group and individual levels, revealing that although all three conditions (self, friend and stranger) looked strikingly similar at the group level, individual differences told a dramatically different story, namely that friend/self activations were tightly correlated throughout the threat-responsive brain, whereas stranger/self activations were not. In light of this, we argue that standard covariate models may be more sensitive in detecting the degree to which a

person processes two conditions in a similar way, whereas conjunction may be sufficient for determining whether the same neural structure is normatively involved in processing stimuli under two different conditions.

The blurring of self and other

Although intra-individual correlations between threat-to-self and threat-to-stranger were few and weak, those between threat-to-self and threat-to-friend were numerous and strong. This suggests that threat-responsive regions of the brain are capable of representing others in a manner that is very similar to the way they represent the self but tend to do so only to the extent that those others are perceived as familiar. This pattern may reflect two distinct forms of empathic responding, one aligned more with a true simulation account and another perhaps aligned more with a 'theory theory' view. For example, the self–friend pattern seems to indicate a more experiential form of empathy, one that is reflected in participant self-reports, where self–friend correlations of subjective valence were also high. This form of empathy seems rooted in the assumption that the other is highly similar to or otherwise aligned with the self, almost as if the boundary between self and other has begun to break down. Indeed, this pattern may reflect a form of *identification* with the other that transcends simple empathic understanding (cf. Decety and Chaminade, 2003). In contrast, threat activation on behalf of strangers is not as yoked to the experience of the self, suggesting that empathic understanding of strangers may be differently mediated. Here, conjunction may offer some important clues. Recall that although self–other correlations differed dramatically as a function of familiarity, group-level conjunction did not. This suggests that although individuals were not using an identification-based strategy during the stranger condition, they did nevertheless activate threat-responsive circuits on the stranger's behalf. This pattern may reflect the supplementary role some 'theory theorists' (Carruthers, 1996) concede to simulation in providing a predictive test ground to assist a theory-based understanding of others. Put another way, although we are less likely to identify with strangers, we can still reference our own experience to make more accurate predictions about what strangers are experiencing.

We should note here that our interpretation of these results is consistent with extant social psychological models of the self in relation to others. For example, Aron *et al.* (cf. Aron and Aron, 1996) have suggested that one way to understand familiarity is through a process whereby the psychological representation of 'self' is expanded to include others, in effect creating the 'breach of individual separateness' that Singer *et al.* (2004) have referred to. From this perspective, the considerable correlations between threat-to-self and threat-to-other activations indicate a form of self–other overlap predicted by self-expansion theory (Aron and Aron, 1996; see also relational self theory, Andersen and Chen, 2002), which argues that the metaphor of including of the other in the self captures an essential aspect of relationship development. This theory presaged much of the development of embodiment approaches to human cognition by suggesting that physical metaphors, such as inclusion of the other in the self, are readily understood by the human mind. Previous work has shown, for example, a tendency for self–other confusion in source recognition memory as a function of familiarity and closeness (Aron and Fraley, 1999; Mashek *et al.*, 2003). Our data add support for self-expansion theory at the neural level, by showing that threat-related neural representations attributable to the self overlap with those attributable to others as a function of familiarity. Indeed, our observation that threat-related activation on behalf of others correlated with the IOS scale (Aron *et al.*, 1992) when the other was a friend, but not when the other was a stranger, lends additional support for this inference.

Still others have proposed models in which 'the boundaries of the self are redrawn' (Brewer and Gardner, 1996, p. 84), suggesting that representations of the self can even extend to the group level (e.g. Brewer and Gardner, 1996) or a mannequin's body, by the manipulation of visual and other sensory systems (Petkova and Ehrsson, 2008, Ramachandran *et al.*, 2011).

Future research and implications for altruism

We speculate that the blurring of self and other representations may be a critical mechanism through which the brain achieves what Tomasello and colleagues (Warneken and Tomasello, 2006; Tomasello and Carpenter, 2007) call 'shared intentionality'. Shared intentionality refers to the sharing of psychological states between individuals and is intimately connected to joint attention, cooperative communication, collaborative action and instructed learning. Tomasello and colleagues suggest in turn that these capacities for shared intentionality, joint attention and cooperative behaviors were critical stepping stones toward the evolution of altruistic behavior in humans.

Converging evidence suggests that altruism evolved in humans under conditions that promoted collaborative behavior. To achieve such behavior, individuals may need to temporarily expand their neural representation of 'self' to include a superordinate unit—a dyad or group. By expanding representations of the self in this way, humans promote the welfare of the group as if they and the group are approximately the same unit. In this sense, altruism is not strictly speaking selfless. Rather, the self is expanded to include other individuals, and the resultant group behaves selfishly, a phenomenon one might call groupishness (cf. Brown, 2000 for a similar definition of groupishness). An intriguing, albeit speculative, hypothesis to draw from this is that altruism motivated by empathy may require some level of overlap in the neural representation of self and other—one that conveys information about this extended self to other brain systems responsible for motivation and action.

CONCLUSION

On the one hand, this study suggests that covariate approaches to investigating self–other overlap in neuroimaging investigations of empathy shed light on patterns and processes that are inaccessible to conjunction analysis. On the other hand, this methodological insight has led to a change in our understanding of empathic responding in the brain. Specifically, the degree to which self-threat-related activation is actually correlated with threat-related activation on behalf of another is dependent on how familiar that other is. Indeed, our results corroborate earlier social psychological suggestions that familiarity involves the inclusion of the other into the self—that from the perspective of the brain, our friends and loved ones are indeed part of who we are.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

CONFLICT OF INTEREST

None declared.

REFERENCES

- Allen, J.P., Porter, M., McFarland, C., McElhane, K.B., Marsh, P. (2007). The relation of attachment security to adolescents' paternal and peer relationships, depression, and externalizing behavior. *Child Development*, 78, 1222–39.
- Andersen, A.M., Chen, S. (2002). The relational self: an interpersonal social-cognitive theory. *Psychological Review*, 109, 619–45.
- Aron, E.N., Aron, A. (1996). Love and expansion of the self: the state of the model. *Personal Relationships*, 3, 45–58.

- Aron, A., Aron, E.N., Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63, 596–612.
- Aron, A., Fraley, B. (1999). Relationship closeness as including other in the self: cognitive underpinnings and measures. *Social Cognition*, 17, 140–60.
- Batson, C.D., Shaw, L.L. (1991). Evidence for altruism: toward a pluralism of prosocial motives. *Psychological Inquiry*, 2, 107–22.
- Berkman, E.T., Lieberman, M.D. (2009). Approaching the bad and avoiding the good: lateral prefrontal cortical asymmetry distinguishes between action and valence. *Journal of Cognitive Neuroscience*, 22, 1970–9.
- Bradley, M., Lang, P.J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavioral Therapy and Experimental Psychology*, 25, 49–59.
- Brewer, M.B., Gardner, W. (1996). Who is this “we”? Levels of collective identity and self representations. *Journal of Personality and Social Psychology*, 71, 83–93.
- Brown, S. (2000). Evolutionary models of music: From sexual selection to group selection. In: Tonneau, F., Thompson, N.S., editors. *Perspectives in Ethology, Volume 13: Behavior, Evolution, and Culture*. NY: Plenum Publishers, pp. 231–82.
- Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J.C., Lenzi, G.L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceeding of the National Academy of Sciences, USA*, 100, 5497–502.
- Carruthers, P. (1996). Simulation and self knowledge: A defense of theory-theory. In: Carruthers, P., Smith, P.K., editors. *Theories of Theories of Mind*. NY: Cambridge University Press, pp. 22–38.
- Cheng, Y., Lin, C., Liu, H.L., et al. (2007). Expertise modulates the perception of pain in others. *Current Biology*, 17, 1708–13.
- Coan, J.A., Allen, J.J.B. (2003). Varieties of emotional experience during voluntary emotional facial expression. *Annals of the New York Academy of Sciences*, 1000, 375–9.
- Coan, J.A., Schaefer, H.S., Davidson, R.J. (2006). Lending a hand: social regulation of the neural response to threat. *Psychological Science*, 17, 1032–9.
- Conneely, K.N., Boehnke, M. (2007). So many correlated tests, so little time! Rapid adjustment of *P* values for multiple correlated tests. *American Journal of Human Genetics*, 81, 1158–68.
- de Vignement, F., Singer, T. (2006). The empathic brain: how when and why? *Trends in Cognitive Science*, 10, 435–41.
- Decety, J. (2011). Dissecting the neural mechanisms mediating empathy. *Emotion Review*, 3, 92–108.
- Decety, J., Chaminade, T. (2003). When the self represents the other: a new cognitive neuroscience view on psychological identification. *Consciousness and Cognition*, 12, 577–96.
- Gallese, V., Goldman, A. (1998). Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Science*, 12, 493–501.
- Hein, G., Singer, T. (2008). I feel how you feel, but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology*, 18, 153–8.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653–70.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17, 825–841.
- Keysers, C., Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Science*, 11, 194–6.
- Lamm, C., Batson, C.D., Decety, J. (2007). The neural substrate of human empathy: effects of perspective taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, 19, 42–58.
- Lamm, C., Meltzoff, A.N., Decety, J. (2010). How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 22, 362–276.
- Mashek, D.J., Aron, A., Boncimo, M. (2003). Confusions of self with close others. *Personality and Social Psychology Bulletin*, 29, 382–92.
- McElhaney, K.B., Antonishak, J., Allen, J.P. (2008). “They like me, they like me not”: Popularity and adolescents’ perceptions of acceptance predicting social functioning over time. *Child Development*, 79, 720–31.
- Nichols, T. (2007). Easythresh_conj—quick method of getting conjunction stats outside of Feat [web page]. Available at: http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/scripts/fsl/easythresh_conj.sh (27 October 2011, date last accessed).
- Ochsner, K.N., Zaki, J., Hanelin, J., et al. (2008). Your pain or mine? Common and distinct neural systems supporting the perception of pain in self and others. *Social Cognitive and Affective Neuroscience*, 3, 144–60.
- Petkova, V.I., Ehrsson, H.H. (2008). If I were you: perceptual illusion of body swapping. *PLoS ONE*, 3(12), e3832. doi:10.1371/journal.pone.0003832.
- Poldrack, R.A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, 2, 67–70.
- Ramachandran, V.S., Krause, B., Case, L.K. (2011). The phantom head. *Perception*, 40, 367–70.
- Singer, T., Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Science*, 1156, 81–96.
- Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303, 1157–62.
- Singer, T., Seymour, B., O’Doherty, J., Stephan, K.E., Dolan, R.J., Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439, 466–9.
- Smith, S.M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17, 143–55.
- Tomasello, M., Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10, 121–5.
- Warneken, F., Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 311, 1301–3.
- Williams, V.S.L., Jones, L.V., Tukey, J.W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics*, 24, 42–69.
- Witt, J.K., Proffitt, D.R., Epstein, W. (2005). Tool use affects perceived distance, but only when you intend to use it. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 880–8.
- Worsley, K.J. (2001). *Functional MRI: An Introduction to Methods*. Oxford: Oxford University Press.