ORIGINAL ARTICLE

Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy

Danilo Bzdok · Leonhard Schilbach · Kai Vogeley · Karla Schneider · Angela R. Laird · Robert Langner · Simon B. Eickhoff

Abstract Morally judicious behavior forms the fabric of human sociality. Here, we sought to investigate neural activity associated with different facets of moral thought. Previous research suggests that the cognitive and emotional sources of moral decisions might be closely related to theory of mind, an abstract-cognitive skill, and empathy, a rapid-emotional skill. That is, moral decisions are thought to crucially refer to other persons' representation of intentions and behavioral outcomes as well as (vicariously experienced) emotional states. We thus hypothesized that moral decisions might be implemented in brain areas engaged in 'theory of mind' and empathy. This assumption was tested by conducting a large-scale activation likelihood estimation (ALE) meta-analysis of neuroimaging studies, which assessed 2,607 peak coordinates from 247 experiments in 1,790 participants. The brain areas that were consistently involved in moral decisions showed more convergence with the ALE analysis targeting theory of mind versus empathy. More specifically, the neurotopographical overlap between morality and empathy disfavors a role of affective sharing during moral decisions. Ultimately, our results provide evidence that the neural network underlying moral decisions is probably domain-global and might be dissociable into cognitive and affective sub-systems.

Keywords Moral cognition · Theory of mind (ToM) · Empathy · Social cognition · Meta-analysis · ALE

D. Bzdok · K. Schneider · R. Langner · S. B. Eickhoff
Department of Psychiatry, Psychotherapy and Psychosomatics, Medical School, RWTH Aachen University, Aachen, Germany

D. Bzdok · R. Langner · S. B. Eickhoff
Institute of Neuroscience and Medicine (INM-2), Research Center Jülich, Jülich, Germany

D. Bzdok · K. Schneider · R. Langner · S. B. Eickhoff
Jülich Aachen Research Alliance (JARA), Translational Brain Medicine, Aachen, Germany

L. Schilbach
Max-Planck-Institute for Neurological Research, Cologne, Germany

L. Schilbach · K. Vogeley
Department of Psychiatry, University of Cologne, Cologne, Germany

K. Vogeley
Institute of Neuroscience and Medicine (INM-3), Research Center Jülich, Jülich, Germany

A. R. Laird
Research Imaging Institute, University of Texas Health Science Center, San Antonio, TX, USA

S. B. Eickhoff
Institute of Clinical Neuroscience and Medical Psychology, Heinrich Heine University, Düsseldorf, Germany

S. B. Eickhoff (✉)
Institut für Neurowissenschaften und Medizin (INM-2), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
e-mail: S.Eickhoff@fz-juelich.de

Introduction

Moral behavior is a building block of human societies and has classically been thought to be based on rational

consideration. Aristotle (fourth century BC/1985), for example, argued that being a "good person" requires reasoning about virtues. Kant's (1785/1993) famous categorical imperative was similarly rational, demanding that one should act according to principles that could also become a general law. More recently, Kohlberg et al. (1983) and Kohlberg (1969) advanced a six-stage developmental model acknowledging children's increasing ability for abstraction and role-taking capacities in moral decisions. In contrast to rational models, the role of emotion in facilitating moral behavior has been less often emphasized (Haidt 2001). Hume (1777/2006) provides an early notable exception, as he believed in a key role of intuition for recognizing morally good and bad decisions, not requiring willful abstract reasoning. Charles Darwin (1874/1997) further argued that moral decisions are mainly influenced by emotional drives, which are rooted in socio-emotional instincts already present in non-human primates.

Notions of rationality and emotionality also serve as explanations in the contemporary neuroscientific literature on the psychological processes underlying moral decisions (henceforth: moral cognition). Results from functional magnetic resonance imaging (fMRI) studies by Greene et al. (2004) and Greene et al. (2001) were interpreted as revealing a dissociation between fast emotional responses and subsequent cognitive modulations in moral cognition. FMRI findings by Moll and Schulkin (2009), Moll et al. (2005a) and Moll et al. (2006), however, were interpreted as revealing a dissociation between group-oriented (i.e., pro-social) and self-oriented (i.e., egoistic) affective drives in moral cognition. Taken together, either abstract-inferential or rapid-emotional processing has been emphasized by most previous accounts explaining moral behavior. Rational explanations assumed that moral behavior arises from a conscious weighing of different rules, norms and situational factors, while emotional explanations emphasized the influence of uncontrolled emotional states rapidly evoked by a given situation (cf. Krebs 2008).

Previous evidence suggests that, the rational and emotional facets of moral cognition are likely related to other social skills: theory of mind (ToM) and empathy. ToM refers to the ability to contemplate other's thoughts, desires, and behavioral dispositions by *abstract inference* (Premack and Woodruff 1978; Frith and Frith 2003). Indeed, accumulating evidence indicates that moral cognition is influenced by whether or not an agent's action is perceived as intentional versus accidental (Knobe 2005; Cushman 2008; Killen et al. 2011; Moran et al. 2011). Empathy, on the other hand, refers to automatically adopting somebody's *emotional state* while maintaining the self–other distinction (Singer and Lamm 2009; Decety and Jackson 2004). In moral decisions, experiencing empathy was shown to alleviate harmful actions towards others (Feshbach and

Feshbach 1969; Zahn-Waxler et al. 1992; Eisenberger 2000). Conversely, the deficient empathy skills in psychopathic populations are believed to contribute to morally inappropriate behavior (Hare 2003; Blair 2007; Soderstrom 2003). Taken together, moral cognition is thought to involve the representation of intentions and outcomes as well as (vicariously experienced) emotional states (Decety et al. 2012; Leslie et al. 2006; Pizarro and Bloom 2003). This assumption is further supported by the observation that ToM and empathic skills precede mature moral reflection in primate evolution (Greene and Haidt 2002) and in ontogeny (Tomasello 2001; Kohlberg et al. 1983; Frith and Frith 2003; Piaget 1932) given that natural evolution tends to modify existing biological systems rather than create new ones from scratch (Jacob 1977; Krebs 2008).

Notably, theoretical accounts as well as empirical evidence suggest that ToM and empathy are partially overlapping psychological constructs. In particular, it has been proposed that embodied representations of affect, which should be relevant for empathic processing, may be further integrated into meta-representational or inferential processing (Keysers and Gazzola 2007; Spengler et al. 2009; De Lange et al. 2008; Mitchell 2005). That is, embodied representation and meta-representation might not constitute two mutually exclusive processes. In other words, more automatic, bottom-up driven mapping and awareness of others emotional states in the context of self–other distinction (i.e., empathy) might be modulated by more controlled, top-down processes involved in attributing mind states (i.e., ToM) (Leiberg and Anders 2006; Pizarro and Bloom 2003; Singer and Lamm 2009). Importantly, ToM and empathy differ in the representational content (mental states versus affect), yet both might be similarly brought about by interaction between top-down and bottom-up processes (Singer and Lamm 2009; Lamm et al. 2007; Spengler et al. 2009; Cheng et al. 2007). In line with the theoretical arguments for a partial overlap between ToM and empathy, ventromedial prefrontal cortex lesions associated with ToM impairments were empirically shown to debilitate elaborate forms of empathic processing, while lateral inferior frontal cortex lesions, which leave ToM skills intact, lead to an impairment of basic forms of empathic processing (Shamay-Tsoory et al. 2009). In sum, converging earlier evidence thus suggests that moral cognition is subserved by partially interrelated ToM and empathy processes.

From a neurobiological perspective, we therefore hypothesized that moral cognition might be subserved by brain areas also related to ToM and empathy, that is, brain networks associated with abstract-inferential and rapid-emotional processing, respectively. Furthermore, we sought to formally investigate to what extent the neurobiological correlates of ToM versus empathy overlap with the neural network associated with moral cognition. Moreover, we

tested whether a common set of brain areas might be implicated in all three of these elaborate social-cognitive skills. We addressed these questions by means of a quantitative meta-analysis on peak coordinates reported in functional neuroimaging studies on moral cognition, ToM, and empathy.

Methods

Data used for the meta-analysis

We searched the PubMed database (http://www.pubmed.org) for fMRI and PET studies that investigated the neural correlates of moral cognition, ToM and empathy. Relevant papers were found by keyword search (search strings: "moral", "morality", "norm", "transgression", "violation", "theory of mind", "mentalizing", "false belief", "perspective taking", "empathy", "empathic", "fMRI", "PET"). Further studies were identified through review articles and reference tracing from the retrieved papers. Please note that in the context of ALE, the term "experiment" usually refers to any single (contrast) analysis on imaging data yielding localization information, while the term "study" usually refers to a scientific publication reporting one or more "experiments" (Laird et al. 2011; Eickhoff and Bzdok 2012). The inclusion criteria comprised full brain coverage as well as absence of pharmacological manipulations, brain lesions or mental/neurological disorders. Additionally, studies were only considered, if they reported results of whole-brain group analyses as coordinates corresponding to a standard reference space (Talairach/Tournoux, MNI). That is, experiments assessing neural effects in a priori defined regions of interest were excluded. We included all eligible neuroimaging studies published up to and including the year 2010. The exhaustive literature search yielded in the moral cognition category a total of 67 experiments reporting 507 activation foci; in the ToM category a total of 68 experiments reporting 724 activation foci, and in the empathy category a total of 112 experiments reporting 1,376 activation foci.

Methodologically equivalent to earlier ALE meta-analyses (Spreng et al. 2009; Lamm et al. 2011; Bzdok et al. 2011; Fusar-Poli et al. 2009; Mar 2011), study selection was grounded on the objective measure whether or not the authors claimed to have isolated brain activity that relates to moral cognition, ToM, or empathy. More specifically, we only included those neuroimaging studies into the "morality" category that required participants to make appropriateness judgments on actions of one individual towards others. In these studies, participants passively viewed or explicitly evaluated mainly textual, sometimes pictorial social scenarios with moral violations/dilemmas.

The target conditions were frequently contrasted with neutral or unpleasant social scenarios (see Supplementary Table 1 for detailed study descriptions). Furthermore, we only included those neuroimaging studies into the "ToM" category that required participants to adopt an intentional stance towards others, that is, predict their thoughts, intentions, and future actions. These studies mostly presented cartoons and short narratives that necessitated understanding the beliefs of the acting characters. The target conditions were usually contrasted with non-social physical stories, which did not necessitate social perspective-taking. Finally, we only included those neuroimaging studies into the "empathy" category that aimed at eliciting the conscious and isomorphic experience of somebody else's affective state. Put differently, in these studies participants were supposed to know and "feel into" what another person was feeling. These studies employed mostly visual, sometimes textual or auditory stimuli that conveyed affect-laden social situations which participants watched passively or evaluated on various dimensions.

Please note that we disregarded studies on empathy for pain because pain, although possessing an affective dimension, is not considered a classic emotion (Izard 1971; Ekman 1982). Rather, it is a bodily sensation mediated by distinct sensory receptors (Craig 2002; Saper 2000), and watching painful scenes does not induce isomorphic vicarious experiences (cf. Singer et al. 2004; Danziger et al. 2009). In particular, looking at a happy person usually elicits a sensation of happiness in the observer, yet watching a person in pain usually does not likewise evoke the physical experience of pain. Furthermore, we disregarded studies in which participants were presented with emotion recognition tasks using static pictures of emotional facial expressions. That is because such tasks are probably too simple to reliably entail sharing others' emotions and maintaining a self–other distinction, both widely regarded as hallmarks of empathy (Singer and Lamm 2009).

Methodological foundation of activation likelihood estimation

The reported coordinates were analyzed for topographic convergence using the revised ALE algorithm for coordinate-based meta-analysis of neuroimaging results (Eickhoff et al. 2009; Turkeltaub et al. 2002; Laird et al. 2009a). The goal of coordinate-based meta-analysis of neuroimaging data is to identify brain areas where the reported foci of activation converge across published experiments. To this end, the meta-analysis determines if the clustering is significantly higher than expected under the null distribution of a random spatial association of results from the considered experiments while acknowledging the spatial uncertainty associated with neuroimaging foci.

As the first step, reported foci were interpreted as centers for 3D Gaussian probability distributions that capture the spatial uncertainty associated with each focus. This uncertainty is mostly a function of between-template (attributable to different normalization strategies and templates across laboratories) and between-subject (due to small sample sizes) variance. In fact, the between-template and between-subject variability are acknowledged based on empirical estimates, the latter being additionally gauged by individual sample size (Eickhoff et al. 2009).

In a second step, the probabilities of all activation foci in a certain experiment were combined for each voxel, yielding a modeled activation (MA) map (Turkeltaub et al. 2011). Voxel-wise ALE scores resulted from the union across these MA maps and quantified the convergence across experiments at each particular location in the brain.

The third and last step distinguished between random and 'true' convergence by comparing the ensuing ALE scores against an empirical null distribution reflecting a random spatial association between the experiments' MA maps (Eickhoff et al. 2012). The within-experiment distribution of foci, however, was regarded to be fixed (Eickhoff et al. 2009). Thus, a random-effects inference was invoked, focusing on the above-chance convergence across different experiments (Eickhoff et al. 2009; Caspers et al. 2010; Kurth et al. 2010). The resulting ALE scores were tested against the earlier calculated 'true' ALE scores and cut off at a cluster-level-corrected threshold of $p < 0.05$. For cluster-level correction, the statistical image of uncorrected voxel-wise p values was first cut off by the cluster-forming threshold. Then, the size of the suprathreshold clusters was compared against a null distribution of cluster sizes derived from simulating 1,000 datasets with the same properties (number of foci, uncertainty, etc.) as the original experiments but random location of foci. The p value associated with each cluster was then given by the chance of observing a cluster of the given size in any particular simulation.

Additional conjunction and difference analyses were conducted to explore how different meta-analyses relate to each other. Conjunction-analyses testing for convergence between different meta-analyses employed inference by the minimum statistic, i.e., computing the intersection of the thresholded Z-maps (Caspers et al. 2010). That is, any voxel determined to be significant by the conjunction analysis constitutes a location in the brain which survived inference corrected for multiple comparisons in each of the individual meta-analyses. Difference analyses calculated the difference between corresponding voxels' ALE scores for two sets of experiments. Then, the experiments contributing to either analysis were pooled and randomly divided into two analogous sets of experiments (Eickhoff et al. 2011). Voxel-wise ALE scores for these two sets

were calculated and subtracted from each other. Repeating this process 10,000 times yielded a null distribution of recorded differences in ALE scores between two sets of experiments. The 'true' difference in ALE scores was then tested against these differences obtained under the null distribution yielding voxel-wise p values for the difference. These resulting non-parametric p values were thresholded at $p < 0.001$. Unfortunately, a statistical method to correct for multiple comparisons when assessing the differences between ALE maps has not yet been established. It should be mentioned, however, that the randomization procedure employed to compute the contrast between ALE-analyses is in itself highly conservative as it estimates the probability for a true difference between the two datasets.

Functional characterization

The converging activation patterns of tasks requiring moral cognition, theory of mind, or empathy were first determined by ALE meta-analysis. The conjunction across all three individual ALE analyses then yielded a computationally derived seed region for functional characterisation through quantitative correspondence with cognitive and experimental descriptions of the BrainMap taxonomy. In fact, BrainMap metadata describe the category of mental processes isolated by the statistical contrast of each experiment stored in the database (http://www.brainmap.org; Laird et al. 2009a). More specifically, behavioral domains (BD) include the main categories cognition, action, perception, emotion, interoception, as well as their subcategories. The respective paradigm classes (PC) categorize the specific task employed (a complete list of BDs and PCs can be found at http://www.brainmap.org/scribe/). We analyzed the behavioral domain and paradigm class metadata of BrainMap experiments associated with seed voxels to determine the frequency of domain 'hits' relative to its likelihood across the entire database. Using a binomial test, Bonferroni-corrected for multiple comparisons, the functional roles of the convergent network as a whole were identified by determining those BDs and PCs that showed a significant over-representation in experiments activating within the seed regions, relative to the entire BrainMap database (Laird et al. 2009b; Eickhoff et al. 2011).

Results

Individual meta-analyses of moral cognition, theory of mind, and empathy

All areas resulting from the ALE meta-analyses were anatomically labeled by reference to probabilistic cytoarchitectonic maps of the human brain using the SPM

Anatomy Toolbox (see Tables 1, 2; Eickhoff et al. 2007; Eickhoff et al. 2005). Meta-analysis of neuroimaging studies related to moral cognition yielded convergent activation in the bilateral ventromedial/frontopolar/dorsomedial prefrontal cortex (vmPFC/FP/dmPFC), precuneus (Prec), temporo-parietal junction (TPJ), posterior cingulate cortex (PCC), as well as the right temporal pole (TP), right middle temporal gyrus (MTG), and left amygdala (AM) (Fig. 1; Table 1). Meta-analysis of studies related to ToM revealed convergence in the bilateral vmPFC/FP/dmPFC,

Table 1 Peaks of activations for the brain areas consistently engaged in fMRI studies on moral cognition, theory of mind, and empathy as revealed by ALE meta-analysis

Macroanatomical location	MNI coordinates		
	x	y	z
ALE meta-analysis of moral cognition			
Ventromedial prefrontal cortex	4	58	−8
Ventromedial prefrontal cortex	−10	42	−18
Frontopolar cortex	0	62	10
Frontopolar cortex	−6	52	18
Dorsomedial prefrontal cortex	0	54	36
Precuneus	0	−56	34
Right temporo-parietal junction (PGa, PGp)	62	−54	16
Left temporo-parietal junction (PGa, PGp)	−48	−58	22
Right temporal pole	54	8	−28
Right middle temporal gyrus	54	−8	−16
Left amygdala	−22	−2	−24
Posterior cingulate cortex	−4	−26	34
ALE meta-analysis of theory of mind			
Ventromedial prefrontal cortex	0	52	−12
Frontopolar cortex	2	58	12
Dorsomedial prefrontal cortex	−8	56	30
Dorsomedial prefrontal cortex	4	58	25
Precuneus	2	−56	30
Right temporo-parietal junction (PGa, PGp)	56	−50	18
Left temporo-parietal junction (PGa, PGp)	−48	−56	24
Right temporal pole	54	−2	−20
Left temporal pole	−54	−2	−24
Right middle temporal gyrus	52	−18	−12
Left middle temporal gyrus	−54	−28	−4
Left middle temporal gyrus	−58	−12	−12
Right posterior superior temporal sulcus	50	−34	0
Left posterior superior temporal sulcus	−58	−44	4
Right inferior frontal gyrus (Area 45)	54	28	6
Left inferior frontal gyrus	−48	30	−12
Right MT/V5	48	−72	8
ALE meta-analysis of empathy			
Dorsomedial prefrontal cortex	2	56	18

Table 1 continued

Macroanatomical location	MNI coordinates		
	x	y	z
Dorsomedial prefrontal cortex	−8	54	34
Right anterior insula	36	22	−8
Left anterior insula	−30	20	4
Right inferior frontal gyrus	50	12	−8
Right inferior frontal gyrus (Area 44)	54	16	20
Right inferior frontal gyrus (Area 45)	50	30	4
Left inferior frontal gyrus	−44	24	−6
Supplementary motor area (Area 6)	−4	18	50
Anterior mid-cingulate cortex	−2	28	20
Rostral anterior cingulate cortex	−4	42	18
Posterior cingulate cortex	−2	−32	28
Right temporo-parietal junction (PGp)	52	−58	22
Left temporo-parietal junction (PGa)	−56	−58	22
Right amygdala	22	−2	−16
Right middle temporal gyrus	54	−8	−16
Right posterior superior temporal sulcus	52	−36	2
Left anterior thalamus	−12	−4	12
Right posterior thalamus	6	−32	2
Right hippocampus (SUB)	26	−26	−12
Midbrain	2	−20	−12
Right pallidum	14	4	0

All peaks are assigned to the most probable brain areas as revealed by the SPM Anatomy Toolbox (Eickhoff et al. 2007; Amunts et al. 2005; Geyer 2004; Caspers et al. 2006; Amunts et al. 1999; Eickhoff et al. 2005)

Prec, TPJ, TP, MTG, posterior superior temporal sulcus (pSTS), and inferior frontal gyrus (IFG), as well as the right MT/V5. Finally, meta-analysis of studies related to empathy yielded convergence in the bilateral dmPFC, supplementary motor area (SMA), rostral anterior cingulate cortex (rACC), anterior mid-cingulate cortex (aMCC; cf. Vogt 2005), PCC, anterior insula (AI), IFG, midbrain, and TPJ, as well as the anterior thalamus on the left, further, the AM, MTG, pSTS, posterior thalamus, hippocampus, and pallidum on the right.

Conjunction analyses

We conducted three conjunction analyses to examine where regions consistently involved in moral cognition converged with regions consistently involved in ToM, empathy, or both (Fig. 2; Table 2). In particular, we computed the conjunction across the individual meta-analysis of moral cognition and the difference between the analyses of ToM and empathy. Employing the difference excluded brain activity shared by both ToM and empathy. The conjunction across the neural network linked to moral cognition and the

Table 2 Conjunction analyses that test for topographical convergence between the individual ALE meta-analyses of moral cognition, theory of mind, and empathy

Macroanatomical location	MNI coordinates		
	x	y	z
Morality ∩ (theory of mind–empathy)			
Ventromedial prefrontal cortex	2	54	−12
Frontopolar cortex	4	60	10
Dorsomedial prefrontal cortex	0	54	30
Right temporo-parietal junction (PGa)	62	−54	14
Left temporo-parietal junction (PGa, PGp)	−50	−58	22
Right middle temporal gyrus	54	−16	−16
Right temporal pole	54	2	−24
Morality ∩ (empathy–theory of mind)			
Dorsomedial prefrontal cortex	−4	50	20
Morality ∩ theory of mind ∩ empathy			
Dorsomedial prefrontal cortex	−5	54	34
Right temporo-parietal junction (PGp)	52	−58	20
Left temporo-parietal junction (PGa)	−54	−58	22
Right middle temporal gyrus	54	−8	−16

All peaks are assigned to the most probable brain areas as revealed by the SPM Anatomy Toolbox (Eickhoff et al. 2007; Caspers et al. 2006; Eickhoff et al. 2005)

activation more robustly linked to ToM versus empathy revealed an overlap in the bilateral vmPFC/FP/dmPFC and TPJ, as well as the right MTG and TP. Conversely, the conjunction across the neural network linked to moral cognition and the activation more robustly linked to empathy versus ToM revealed an overlap in the dmPFC. Besides results from this conjunction analysis, it is noteworthy that the left AM was found in the analysis of moral cognition-related brain activity, while the right AM was found in the analysis of empathy-related brain activity. Moreover, the PCC showed significant convergence in both these individual meta-analyses at adjacent, yet non-overlapping locations. Specifically, convergence in the PCC was located slightly more rostro-dorsally in the ALE on moral cognition compared to the ALE on empathy (Fig. 1; Table 1). Finally, the conjunction analysis across the results of all three individual meta-analyses on moral cognition, ToM, and empathy yielded convergence in the dmPFC, right MTG, and bilateral TPJ. Additionally, we provide a list of those papers included in the present meta-analysis, which gave rise to the four converging clusters (Supplementary Table 2).

Functional characterization

Using the metadata of the BrainMap database for functional characterization as described above, we found that the
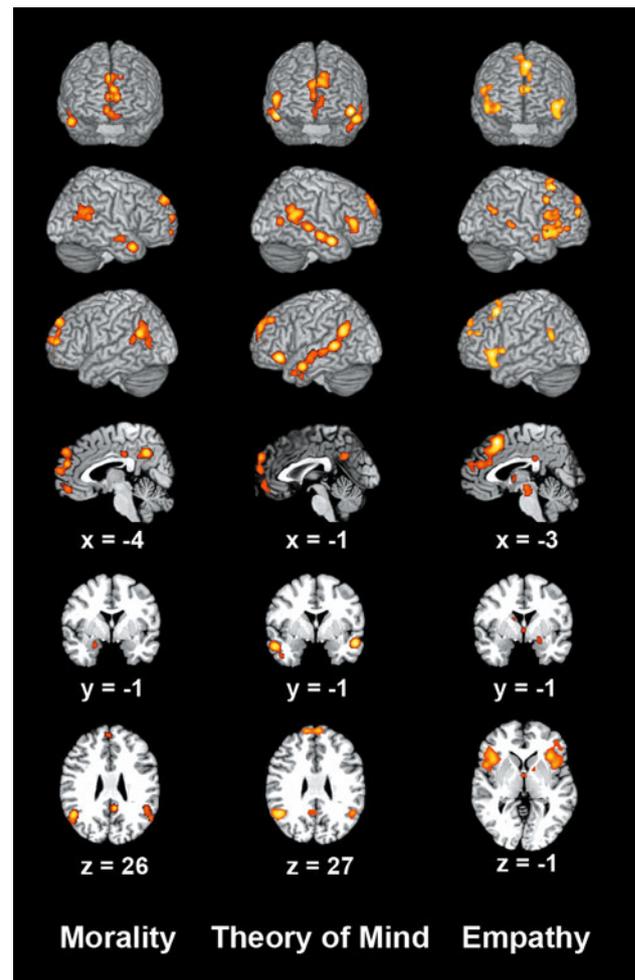

Fig. 1 ALE meta-analysis of neuroimaging studies on moral cognition, theory of mind, and empathy. Significant meta-analysis results displayed on frontal, right, and left surface view as well as sagittal, coronal, and axial sections of the MNI single-subject template. Coordinates in MNI space. All results were significant at a cluster-forming threshold of $p < 0.05$ and an extent threshold of $k = 10$ voxels (to exclude presumably incidental results)

common network observed across all three task domains was selectively associated with tasks tapping ToM, semantic processing, imagination, and social cognition, as well as episodic and explicit memory (Fig. 3). The network thus appears to be exclusively involved in higher-level cognitive processing.

Discussion

Classic accounts of the psychology of moral cognition can be broadly divided into views that emphasize the involvement of either 'rational' or 'emotional' processes. In this study, we revisited this distinction by determining the consistent overlap between brain activation patterns
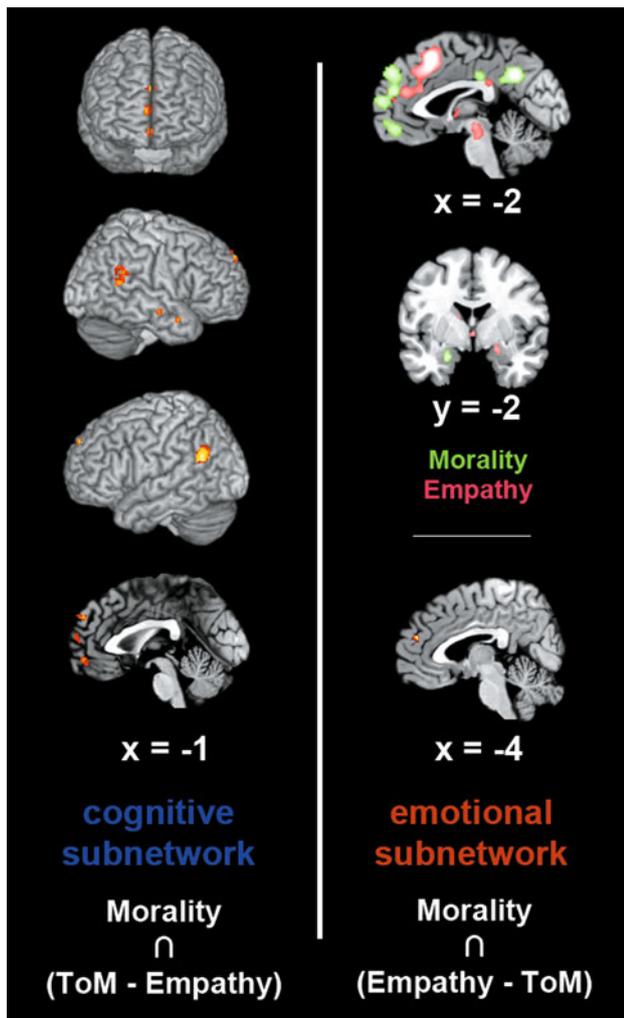
Fig. 2 Conjunction analyses for topographical convergence across brain activity related to moral cognition and theory of mind (ToM) or empathy. *Left panel* overlapping activation patterns between the meta-analysis on moral cognition and the difference analysis between ToM and empathy (cluster-forming threshold: $p < 0.05$). *Right bottom panel* overlapping activation patterns between the meta-analysis on moral cognition and the difference analysis between empathy and ToM (cluster-forming threshold: $p < 0.05$). *Right top panel* sagittal and coronal slices of juxtaposed results from the meta-analyses on moral cognition (*green*) and empathy (*red*) to highlight similar convergence in the posterior cingulate cortex and amygdala (extent threshold: $k = 10$ voxels to exclude presumably incidental results). Coordinates in MNI space

reported in the neuroscientific literature on moral cognition with those of ToM (assumed to be more rational) and empathy (assumed to be more emotional) using coordinate-based ALE meta-analysis. By doing so, we demonstrated that moral cognition, indeed, recruits brain areas that are also involved in abstract-inferential (ToM) and rapid-emotional (empathy) social cognition. Furthermore, TPJ, mPFC, and MTG emerged as potential nodes of a network common to moral cognition, ToM, and empathy.

Individual analyses on the neural correlates of moral cognition, theory of mind, and empathy

To our knowledge, we here conducted the first quantitative meta-analytic assessment of the neural network engaged in moral cognition, that is, reflection of the social appropriateness of people's actions. The obtained pattern of converged brain activation is in very good agreement with qualitative reviews of fMRI studies on moral cognition (Moll et al. 2005b; Greene and Haidt 2002). Concurrently, dysfunction in prefrontal, temporal and amygdalar regions is discussed as linked to psychopathy and anti-social behavior based on neuropsychological, lesion, and fMRI studies (Blair 2007; Anderson et al. 1999; Mendez et al. 2005).

The results of the meta-analysis of ToM were consistent with earlier quantitative analyses (Spreng et al. 2009; Mar 2011) of neuroimaging studies, in which participants attributed mental states to others to predict or explain their behavior. Put differently, the obtained network is likely to be implicated in the recognition and processing of others' *mental states* (Amodio and Frith 2006; Gallagher and Frith 2003; Schilbach et al. 2010).

The meta-analysis on empathy across various affective modalities excluding pain was based on neuroimaging studies in which participants understood and vicariously shared the emotional experience of others. This analysis revealed the aMCC extending into the SMA and the AI extending into the IFG as the most prominent points of convergence. This is in line with a recent image-based meta-analysis on empathy for pain (Lamm et al. 2011). The present analysis on non-pain empathy, however, additionally revealed activation in the AM, rACC, and PCC. Consequently, empathy for pain- and non-pain-related affect appears to be implemented by an overlapping network that might recruit supplementary areas/networks depending on the specific affective modality with which participants are to empathize. In other words, the observed network is likely to be involved in vicariously experiencing others' *affective states* (Singer et al. 2004; Wicker et al. 2003).

It is noteworthy that all brain areas revealed by the meta-analysis on moral cognition also converged significantly in the analyses on either ToM or empathy. It is thus tempting to speculate that moral cognition might rely on remodeling mental states and processing affective states of other people. Conversely, some of the brain areas, which show significant convergence in the ALE on ToM and empathy, were not significant in the ALE on moral cognition. This suggests that moral cognition might be realized by specific subsets, rather than the entirety of the neural correlates of ToM and empathy.
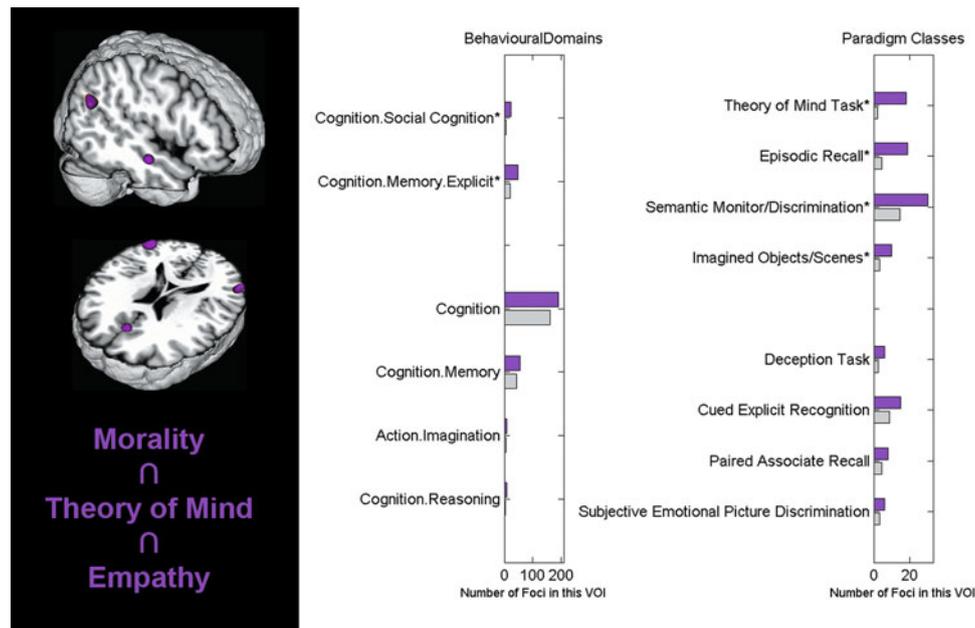
Fig. 3 Functional characterization of the core-network implicated in moral cognition, theory of mind (ToM), and empathy. *Left* neural network consistently activated across individual meta-analyses on moral cognition, ToM, and empathy (extent threshold: $k = 10$ voxels to exclude presumably incidental results). Images were rendered using Mango (multi-image analysis GUI, Research Imaging Institute, San Antonio, Texas, USA; http://ric.uthscsa.edu/mango/). *Right* functional characterization of the convergent network across all three tasks by BrainMap metadata. The *purple bars* denote the number of foci for that particular metadata class within the seed network. The *grey bars* represent the number of foci that would be expected to hit the particular seed network if all foci with the respective class were randomly distributed throughout the cerebral cortex. That is, the *grey bars* denote the by-chance frequency of that particular label given the size of the cluster. All shown taxonomic classes reached significance according to a binomial test ($p < 0.05$). *Asterisks* denote classes that survived the Bonferroni correction for multiple comparisons

Converging neural correlates across moral cognition and theory of mind

Brain activity during moral cognition and ToM overlapped in the bilateral vmPFC/FP/dmPFC and TPJ, as well as the right TP and MTG. This extensive convergence indicates that moral cognition and ToM engage a highly similar neural network, which, in turn, entices to speculate about a close relationship between these two psychological processes. More specifically, increased activity along the dorso-ventral axis of the medial prefrontal cortex is heterogeneously discussed to reflect cognitive versus affective processes (Shamay-Tsoory et al. 2006), controlled/explicit versus automatic/implicit social cognition (Lieberman 2007; Forbes and Grafman 2010), goal versus outcome pathways (Krueger et al. 2009), dissimilarity versus similarity to self (Mitchell et al. 2006) as well as other-focus versus self-focus (Van Overwalle 2009). Apart from that, the spatially largest convergence across moral cognition- and ToM-related brain activity in the bilateral posterior temporal lobe/angular gyrus might be surprising given the divergent interpretation in the literature. That is, activation in this cortical region is conventionally interpreted as "posterior superior temporal sulcus" in the morality literature (Moll et al. 2005b; Greene and

Haidt 2002) and as "temporo-parietal junction" in the ToM literature (Decety and Lamm 2007; Van Overwalle 2009; Saxe and Kanwisher 2003). This convergent activation in the ALE on moral cognition is, however, more accurately located to the TPJ, rather than to the pSTS (cf. Raine and Yang 2006; Binder et al. 2009). The potentially inconsistent neuroanatomical labeling might have disadvantageously affected discussion of this brain area in previous neuroimaging studies on moral cognition.

The engagement of ToM-associated areas during moral cognition has also been addressed in several recent fMRI studies by Young, Saxe, and colleagues. These authors proposed that during moral cognition, the dmPFC might process belief valence, while the TPJ and Prec might encode and integrate beliefs with other relevant features (Young and Saxe 2008). Especially, brain activity in the right TPJ was advocated to reflect belief processing during moral cognition. This argument was based on interaction effects with moral reasoning (Young et al. 2007), correlation with the participants' self-reported tendency for acknowledging belief information (Young and Saxe 2009), and a significantly reduced impact of intentions after transient TPJ disruption (Young et al. 2010b). In line with our results, these fMRI studies suggest that moral cognition

might involve reconstructing attributes and intentions that we apply to others and vice versa.

However, earlier neuroimaging evidence concerning the likely implication of the ToM network in moral cognition has two weak points. First, the interpretations of relevant fMRI studies were largely driven by reverse inference, i.e., identifying psychological processes from mere topography of brain activity (Poldrack 2006). That is, rather than investigating moral cognition tasks alone, the neural networks of moral cognition and ToM should have been compared directly by independent conditions involving either task. Second, the relevant fMRI studies crucially hinge on the repeatedly criticized functional localizer technique, which rests on strong a priori hypotheses (Friston et al. 2006; Mitchell 2008). The present ALE meta-analysis overcomes these two limitations by a *largely hypothesis-free* assessment of two *independent* pools of numerous whole-brain neuroimaging studies, providing strong evidence for the high convergence across the neural networks associated with moral cognition and ToM.

Converging neural correlates across moral cognition and empathy

Brain activity related to both moral cognition and non-pain empathy converged significantly in an area of the dmPFC, which was not revealed by the ALE on ToM. Converging activation of the dmPFC may suggest an implication of this highly associative cortical area in more complex social-emotional processing. In line with this interpretation, a recent fMRI study identified a similar brain location as highly selective for processing guilt (Wagner et al. 2012), an emotion closely related to moral and social transgression (Tangney et al. 2007). Moreover, the dmPFC has consistently been related to the (possibly interwoven) reflection of own and simulation of others' mind states (Lamm et al. 2007; Jenkins and Mitchell 2010; Mitchell et al. 2006; D'Argembeau et al. 2007). Nevertheless, we feel that it might be currently unwarranted to confer precise functions to circumscribed parts of the dmPFC (cf. above), given the danger of reverse inference when deducing mental functions or states from regional activation patterns (Poldrack 2006). We therefore cautiously conclude that the observed convergence in the dmPFC probably reflects an unidentified, yet to be characterized, higher-level neural process that is related to affective and social processing.

It is interesting to note that the AM was significantly involved in the individual ALEs on morality and non-pain empathy, although in contralateral hemispheres. This brain region is thought to automatically extract biological significance from the environment (Ball et al. 2007; Sander et al. 2003; Bzdok et al. 2011; Müller et al. 2011). In particular, AM activity typically increases in the left

hemisphere in elaborate social-cognitive processes and in the right hemisphere in automatic, basic emotional processes (Phelps et al. 2001; Bzdok et al. 2012; Markowitsch 1998; Gläscher and Adolphs 2003). This lateralization pattern thus seems well in line with the consistent engagement of the left AM in moral cognition and right AM in non-pain empathy. Moreover, activity in the PCC was found in adjacent, yet non-overlapping, locations during moral cognition and non-pain empathy. The PCC is thought important for the modality-independent retrieval of autobiographical memories and their integration with current emotional states (Fink et al. 1996; Maddock et al. 2001; Maddock 1999; Schilbach et al. 2008a). It might therefore be speculated that moral judgments and empathic processing could both rely on the integration on past experiences. Given that both PCC clusters do not overlap, however, it remains to be investigated whether (a) moral cognition and non-pain empathy engage distinct regions in the PCC, (b) the observed topographic pattern is purely incidental given the limited spatial resolution of meta-analyses, or (c) the differences in PPC activation reflect differences in stimulus-material and hence autobiographic associations.

Furthermore, it is important to note that the AI/IFG and aMCC/SMA were revealed as the most significant points of convergence in the ALE on non-pain empathy but did not show any overlap with the ALE on morality. In particular, the AI/IFG and aMCC/SMA form a network that is widely believed to represent one's own and others' emotional states regardless of the actual affective or sensory modality (Lamm et al. 2011; Wicker et al. 2003; Singer et al. 2004; Lamm and Singer 2010; Fan et al. 2011). Moreover, this network, especially the anterior insula, is not only implicated in meta-representation of emotional states but also in interoceptive awareness (Craig 2002, 2009; Kurth et al. 2010). Interoception-related meta-representation of emotion has thus been suggested to underlie the concomitant involvement of the aMCC/SMA and AI/IFG in neuroimaging studies on empathy as reflecting affective sharing (Singer and Lamm 2009; Fan et al. 2011).

In summary, we draw three conclusions from these observations. First, affective sharing, one core aspect of empathy, is unlikely to be involved in moral judgments, given the lack of consistent involvement of the AI/IFG or aMCC/SMA in paradigms probing the latter. Second, general affective processes might play a role in both moral judgments and empathy, given that the respective meta-analyses individually revealed the dmPFC (direct overlap) and the amygdala (oppositely lateralized). Third, the neural correlates of moral judgments are much closer related to the neural correlates of ToM than to those of empathy, as evidenced by the quantity of overlap in the conjunction analyses.

Converging neural correlates across all individual analyses

Moral cognition-, ToM-, and empathy-related brain activity converged in the bilateral TPJ, dmPFC, and right MTG, which therefore form a common network potentially involved in social-cognitive processes. In line with this, metadata profiling demonstrates solid associations of this network with neuroimaging studies related to ToM, explicit memory retrieval, language, and imagination of objects/scenes. Intriguingly, these four seemingly disparate psychological categories summarize what sets humans probably apart from non-human primates (Tomasello 2001; Frith and Frith 2010). They might functionally converge in the reciprocal relationship between the allocentric and egocentric perspective, instructed by self-reflection, social knowledge, and memories of past experiences (of social interactions). In particular, autobiographical memory supplies numerous building blocks of social semantic knowledge (Bar 2007; Binder et al. 2009). These isolated conceptual scripts may be reassembled to enable the forecasting of future events (Tulving 1983, 1985; Schacter et al. 2007). Similar brain mechanisms in remembering past episodes and envisioning future circumstances is supported by their engagement of identical brain areas, as evidenced by a quantitative meta-analysis (Spreng et al. 2009). Moreover, retrograde amnesic patients were reported to be impaired not only in prospection but also in imagining novel experience (Hassabis et al. 2007). These findings suggest a single neural network for mentally constructing plausible semantic scenarios of detached situations regardless of temporal orientation (Buckner and Carroll 2007; Hassabis and Maguire 2007). Indeed, construction of detached probabilistic scenes has been argued to influence ongoing decision making by estimating outcomes of behavioral choices (Boyer 2008; Suddendorf and Corballis 2007; Schilbach et al. 2008b). Taken together, moral cognition, ToM, and empathy jointly engage a network that might be involved in the automated prediction of social events that modulate behavior.

Relation to clinical research

Consistent with the demonstrated functional dissociation between cognitive and affective subsystems of the neural network related to moral cognition observed in our study, frontotemporal dementia has been reported to impair personal but not impersonal moral reasoning (Mendez et al. 2005). A cognitive-affective dissociation of moral cognition is also supported clinically by the psychopathic population's immoral behavior in everyday life despite excellent moral reasoning skills (Cleckley 1941; Hare 1993). Apart from that, neither Greene's nor Moll's concept can

exhaustively explain why vmPFC patients demonstrated a rationally biased approach to solving moral dilemmas (Koenigs et al. 2007; Moretto et al. 2010), yet, an emotionally biased approach to moral cognition in an economic game (Koenigs and Tranel 2007). Given the considerable neural commonalities of moral cognition and ToM tasks as revealed by the present analysis, contradictory findings of vmPFC patients dealing with moral dilemmas and economic games (Greene 2007; Koenigs and Tranel 2007; Koenigs et al. 2007) might resolve when meticulously probing ToM capabilities in future lesion studies.

Limitations

Several limitations of our study should be addressed. Meta-analyses are necessarily based on the available literature and hence may be affected by the potential publication-bias disfavoring null results (Rosenthal 1979). This caveat especially applies to the functional characterization of converging activation foci using the BrainMap database, as this database only contains about 21% of the published neuroimaging studies and its content therefore does not constitute a strictly representative sample. Furthermore, a part of the included neuroimaging studies on moral cognition might suffer from limited ecological validity. That is, the experimental tasks used might only partially involve the mental processes guiding real-life moral behavior (cf. Cima et al. 2010; Knutson et al. 2010; Young et al. 2010a). In particular, employing overly artificial moral scenarios (e.g., trolley dilemma), on top of the inherent limitations of neuroimaging paradigms, could have systematically overestimated cognitive versus emotional processes. This caveat might have contributed to the similar results in the meta-analyses on moral cognition and on ToM (rather than on empathy), given that emotion processing is thought to play a paramount role in real-world moral cognition (Krebs 2008; Tangney et al. 2007; Haidt 2001). Future neuroimaging studies should therefore strive for using more realistic moral scenarios to minimize the risk of investigating the neural correlates of "in vitro moral cognition" (cf. Schilbach et al. 2012; Schilbach 2010).

Conclusions

It is a topic of intense debate whether social cognition is subserved by a unitary specialized module or by a set of general-purpose mental operations (Mitchell 2006; Spreng et al. 2009; Bzdok et al. 2012; Van Overwalle 2011). The present large-scale meta-analysis provides evidence for a domain-global view of moral cognition, rather than for a distinct moral module (Hauser 2006), by showing its

functional integration of distributed brain networks. More specifically, we parsed the neural correlates of moral cognition by reference to a *socio-cognitive* framework, exemplified by ToM cognition, and a *socio-affective* framework, exemplified by empathy. Ultimately, our results support the notion that moral reasoning is related to both seeing things from other persons' points of view and to grasping others' feelings (Piaget 1932; Tomasello 2001; Decety et al. 2012).

References

Amodio DM, Frith CD (2006) Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7(4):268–277. doi:10.1038/nrn1884

Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HB, Zilles K (1999) Broca's region revisited: cytoarchitecture and inter-subject variability. J Comp Neurol 412(2):319–341. doi:10.1002/(SICI)1096-9861(19990920)412:2<319:AID-CNE10>3.0.CO;2-7

Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, Habel U, Schneider F, Zilles K (2005) Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. Anat Embryol (Berl) 210(5–6):343–352. doi:10.1007/s00429-005-0025-5

Anderson SW, Bechara A, Damasio H, Tranel D, Damasio AR (1999) Impairment of social and moral behavior related to early damage in human prefrontal cortex. Nat Neurosci 2(11):1032–1037. doi:10.1038/14833

Aristotle (1985) Nicomachean Ethics. Hackett, Indianapolis

Ball T, Rahm B, Eickhoff SB, Schulze-Bonhage A, Speck O, Mutschler I (2007) Response properties of human amygdala subregions: evidence based on functional MRI combined with probabilistic anatomical maps. PLoS One 2(3):e307. doi:10.1371/journal.pone.0000307

Bar M (2007) The proactive brain: using analogies and associations to generate predictions. Trends Cogn Sci 11(7):280–289. doi:10.1016/j.tics.2007.05.005

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex 19(12):2767–2796. doi:10.1093/cercor/bhp055

Blair RJ (2007) The amygdala and ventromedial prefrontal cortex in morality and psychopathy. Trends Cogn Sci 11(9):387–392. doi:10.1016/j.tics.2007.07.003

Boyer P (2008) Evolutionary economics of mental time travel? Trends Cogn Sci 12(6):219–224. doi:10.1016/j.tics.2008.03.003

Buckner RL, Carroll DC (2007) Self-projection and the brain. Trends Cogn Sci 11(2):49–57. doi:10.1016/j.tics.2006.11.004

Bzdok D, Langner R, Caspers S, Kurth F, Habel U, Zilles K, Laird A, Eickhoff SB (2011) ALE meta-analysis on facial judgments of trustworthiness and attractiveness. Brain Struct Funct 215(3–4):209–223. doi:10.1007/s00429-010-0287-4

Bzdok D, Langner R, Hoffstaedter F, Turetsky B, Zilles K, Eickhoff SB (2012) The modular neuroarchitecture of social judgments on faces. Cereb Cortex (in press)

Caspers S, Geyer S, Schleicher A, Mohlberg H, Amunts K, Zilles K (2006) The human inferior parietal cortex: cytoarchitectonic parcellation and interindividual variability. Neuroimage 33:430–448

Caspers S, Zilles K, Laird AR, Eickhoff SB (2010) ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50(3):1148–1167. doi:10.1016/j.neuroimage.2009.12.112

Cheng Y, Lin CP, Liu HL, Hsu YY, Lim KE, Hung D, Decety J (2007) Expertise modulates the perception of pain in others. Curr Biol 17(19):1708–1713. doi:10.1016/j.cub.2007.09.020

Cima M, Tonnaer F, Hauser MD (2010) Psychopaths know right from wrong but don't care. Soc Cogn Affect Neurosci 5:59–67

Cleckley H (1941) The mask of sanity: an attempt to clarify some issues about the so-called psychopathic personality. CV Mosby, St. Louis

Craig AD (2002) How do you feel? Interoception: the sense of the physiological condition of the body. Nat Rev Neurosci 3(8):655–666. doi:10.1038/Nrn894

Craig AD (2009) How do you feel—now? The anterior insula and human awareness. Nat Rev Neurosci 10(1):59–70. doi:10.1038/Nrn2555

Cushman F (2008) Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. Cognition 108:353–380

D'Argembeau A, Ruby P, Collette F, Degueldre C, Balteau E, Luxen A, Maquet P, Salmon E (2007) Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. J Cogn Neurosci 19(6):935–944. doi:10.1162/jocn.2007.19.6.935

Danziger N, Faillenot I, Peyron R (2009) Can we share a pain we never felt? Neural correlates of empathy in patients with congenital insensitivity to pain. Neuron 61(2):203–212. doi:10.1016/j.neuron.2008.11.023

Darwin C (1874/1997) The descent of man. Prometheus Books, New York

De Lange FP, Spronk M, Willems RM, Toni I, Bekkering H (2008) Complementary systems for understanding action intentions. Curr Biol 18:454–457

Decety J, Jackson PL (2004) The functional architecture of human empathy. Behav Cogn Neurosci Rev 3(2):71–100. doi:10.1177/1534582304267187

Decety J, Lamm C (2007) The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist 13(6):580–593. doi:10.1177/1073858407304654

Decety J, Michalska KJ, Kinzler KD (2012) The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. Cereb Cortex. doi:10.1093/cercor/bhr111

Eickhoff SB, Bzdok D (2012) Meta-analyses in basic and clinical neuroscience: State of the art and perspective. In: Ulmer S, Jansen O (eds) fMRI—Basics and Clinical Applications, 2nd edn. Springer, Berlin (in press)

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25(4):1325–1335. doi:10.1016/j.neuroimage.2004.12.034

Eickhoff SB, Paus T, Caspers S, Grosbras MH, Evans AC, Zilles K, Amunts K (2007) Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage 36(3):511–521. doi:10.1016/j.neuroimage.2007.03.060

Eickhoff SB, Laird AR, Grefkes C, Wang LE, Zilles K, Fox PT (2009) Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum Brain Mapp 30(9):2907–2926. doi:10.1002/hbm.20718

Eickhoff SB, Bzdok D, Laird AR, Roski C, Caspers S, Zilles K, Fox PT (2011) Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. Neuroimage 57(3):938–949. doi:10.1016/j.neuroimage.2011.05.021

Eickhoff SB, Bzdok D, Laird AR, Kurth F, Fox PT (2012) Activation likelihood estimation meta-analysis revisited. Neuroimage. doi:10.1016/j.neuroimage.2011.09.017

Eisenberger N (2000) Emotion, regulation, and moral development. Ann Rev Psychol 51:665–697

Ekman P (1982) Emotion in the human face. Cambridge University Press, New York

Fan Y, Duncan NW, de Greck M, Northoff G (2011) Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. Neurosci Biobehav Rev 35(3):903–911. doi:10.1016/j.neubiorev.2010.10.009

Feshbach ND, Feshbach S (1969) The relationship between empathy and aggression in two age groups. Dev Psychol 1:102–107

Fink GR, Markowitsch HJ, Reinkemeier M, Bruckbauer T, Kessler J, Heiss WD (1996) Cerebral representation of one's own past: neural networks involved in autobiographical memory. J Neurosci 16(13):4275–4282

Forbes CE, Grafman J (2010) The role of the human prefrontal cortex in social cognition and moral judgment. Annu Rev Neurosci 33:299–324

Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. Neuroimage 30(4):1077–1087. doi:10.1016/j.neuroimage.2005.08.012

Frith U, Frith CD (2003) Development and neurophysiology of mentalizing. Philos Trans R Soc Lond B Biol Sci 358(1431):459–473. doi:10.1098/rstb.2002.1218

Frith U, Frith C (2010) The social brain: allowing humans to boldly go where no other species has been. Philos Trans R Soc Lond B Biol Sci 365(1537):165–176. doi:10.1098/rstb.2009.0160

Fusar-Poli P, Placentino A, Carletti F, Landi P, Allen P, Surguladze S, Benedetti F, Abbamonte M, Gasparotti R, Barale F, Perez J, McGuire P, Politi P (2009) Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. J Psychiatry Neurosci 34(6):418–432

Gallagher HL, Frith CD (2003) Functional imaging of 'theory of mind'. Trends Cogn Sci 7(2):77–83 pii:S1364661302000256

Geyer S (2004) The microstructural border between the motor and the cognitive domain in the human cerebral cortex. Adv Anat Embryol Cell Biol 174(I–VIII):1–89

Gläscher J, Adolphs R (2003) Processing of the arousal of subliminal and supraliminal emotional stimuli by the human amygdala. J Neurosci 23(32):10274–10282

Greene JD (2007) Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. Trends Cogn Sci 11(8):322–323. doi:10.1016/j.tics.2007.06.004 (author reply 323–324)

Greene JD, Haidt J (2002) How (and where) does moral judgment work? Trends Cogn Sci 6(12):517–523 pii:S1364661302020119

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. Science 293(5537):2105–2108. doi:10.1126/science.1062872

Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. Neuron 44(2):389–400. doi:10.1016/j.neuron.2004.09.027

Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol Rev 108(4):814–834. doi:10.1037/0033-295x.108.4.814

Hare RD (1993) Without conscience: the disturbing world of the psychopath among us. Pocket Books, New York

Hare RD (2003) The hare psychopathy checklist—revised, 2nd edn. Multi-Health Systems, Toronto

Hassabis D, Maguire EA (2007) Deconstructing episodic memory with construction. Trends Cogn Sci 11(7):299–306. doi:10.1016/j.tics.2007.05.001

Hassabis D, Kumaran D, Vann SD, Maguire EA (2007) Patients with hippocampal amnesia cannot imagine new experiences. Proc Natl Acad Sci USA 104(5):1726–1731. doi:10.1073/pnas.0610561104

Hauser MD (2006) The liver and the moral organ. Soc Cogn Affect Neurosci 1(3):214–220. doi:10.1093/scan/nsl026

Hume D (1777/2006) An inquiry concerning the principles of morals. Cosimo Classics, New York

Izard CE (1971) The face of emotion. Appleton-Century-Crofts, New York

Jacob F (1977) Evolution and tinkering. Science 196(4295):1161–1166

Jenkins AC, Mitchell JP (2010) Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. Cereb Cortex 20(2):404–410. doi:10.1093/cercor/bhp109

Kant I (1785/1993) Grounding for the Metaphysics of Morals. Hackett Publishing Co, Indianapolis

Keysers C, Gazzola V (2007) Integrating simulation and theory of mind: from self to social cognition. Trends Cogn Sci 11(5):194–196. doi:10.1016/j.tics.2007.02.002

Killen M, Mulvey KL, Richardson C, Jampol N, Woodward A (2011) The accidental transgressor: morally–relevant theory of mind. Cognition 119:197–215

Knobe J (2005) Theory of mind and moral cognition: exploring the connections. Trends Cogn Sci 9(8):357–359

Knutson KM, Krueger F, Koenigs M, Hawley A, Escobedo JR, Vasudeva V, Adolphs R, Grafman J (2010) Behavioral norms for condensed moral vignettes. Soc Cogn Affect Neurosci 5(4):378–384. doi:10.1093/scan/nsq005

Koenigs M, Tranel D (2007) Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. J Neurosci 27(4):951–956. doi:10.1523/JNEUROSCI.4606-06.2007

Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A (2007) Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 446(7138):908–911. doi:10.1038/nature05631

Kohlberg L (1969) Stage and sequence: the cognitive-developmental approach to socialization. In: Goslin DA (ed) Handbook of socialization theory and research. Rand McNally, Chicago, pp 347–480

Kohlberg L, Levine C, Hewer A (1983) Moral stages: A current formulation and a response to critics. Karger, Basel

Krebs DL (2008) Morality: an evolutionary account. Perspect Psychol Sci 3(3):149–172

Krueger F, Barbey AK, Grafman J (2009) The medial prefrontal cortex mediates social event knowledge. Trends Cogn Sci 13(3):103–109. doi:10.1016/j.tics.2008.12.005

Kurth F, Zilles K, Fox PT, Laird AR, Eickhoff SB (2010) A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. Brain Struct Funct 214(5–6):519–534. doi:10.1007/s00429-010-0255-z

Laird AR, Eickhoff SB, Kurth F, Fox PM, Uecker AM, Turner JA, Robinson JL, Lancaster JL, Fox PT (2009a) ALE meta-analysis workflows via the Brainmap database: progress towards a probabilistic functional brain atlas. Front Neuroinformatics 3:23. doi:10.3389/neuro.11.023.2009

Laird AR, Eickhoff SB, Li K, Robin DA, Glahn DC, Fox PT (2009b) Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling.

J Neurosci 29(46):14496–14505. doi:10.1523/JNEUROSCI.4004-09.2009

Laird AR, Eickhoff SB, Fox PM, Uecker AM, Ray KL, Saenz JJ Jr, McKay DR, Bzdok D, Laird RW, Robinson JL, Turner JA, Turkeltaub PE, Lancaster JL, Fox PT (2011) The BrainMap Strategy for Standardization, Sharing, and Meta-Analysis of Neuroimaging Data. BMC Res Notes 4(1):349. doi:10.1186/1756-0500-4-349

Lamm C, Singer T (2010) The role of anterior insular cortex in social emotions. Brain Struct Funct 214(5):579–591. doi:10.1007/s00429-010-0251-3

Lamm C, Batson CD, Decety J (2007) The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. J Cogn Neurosci 19(1):42–58

Lamm C, Decety J, Singer T (2011) Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54(3):2492–2502. doi:10.1016/j.neuroimage.2010.10.014

Leiberg S, Anders S (2006) The multiple facets of empathy: a survey of theory and evidence. Prog Brain Res 156:419–440

Leslie AM, Knobe J, Cohen A (2006) Acting intentionally and the side-effect effect: theory of mind and moral judgment. Psychol Sci 17:421–427

Lieberman MD (2007) Social cognitive neuroscience: a review of core processes. Annu Rev Psychol 58:259–289. doi:10.1146/annurev.psych.58.110405.085654

Maddock RJ (1999) The retrosplenial cortex and emotion: new insights from functional neuroimaging of the human brain. Trends Neurosci 22(7):310–316 pii:S0166223698013745

Maddock RJ, Garrett AS, Buonocore MH (2001) Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. Neuroscience 104(3):667–676. doi:pii:S0306-4522(01)00108-7

Mar RA (2011) The neural bases of social cognition and story comprehension. Annu Rev Psychol 62:103–134. doi:10.1146/annurev-psych-120709-145406

Markowitsch HJ (1998) Differential contribution of right and left amygdala to affective information processing. Behav Neurol 11(4):233–244

Mendez MF, Anderson E, Shapira JS (2005) An investigation of moral judgement in frontotemporal dementia. Cogn Behav Neurol 18(4):193–197 pii:00146965-200512000-00002

Mitchell JP (2005) The false dichotomy between simulation and theory–theory: the argument's error. Trends Cogn Sci 9(8):363–364. doi:10.1016/j.tics.2005.06.010 (author reply 364)

Mitchell JP (2006) Mentalizing and Marr: an information processing approach to the study of social cognition. Brain Res 1079(1):66–75. doi:10.1016/j.brainres.2005.12.113

Mitchell JP (2008) Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb Cortex 18(2):262–271. doi:10.1093/cercor/bhm051

Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50(4):655–663. doi:10.1016/j.neuron.2006.03.040

Moll J, Schulkin J (2009) Social attachment and aversion in human moral cognition. Neurosci Biobehav Rev 33(3):456–465. doi:10.1016/j.neubiorev.2008.12.001

Moll J, de Oliveira-Souza R, Moll FT, Ignacio FA, Bramati IE, Caparelli-Daquer EM, Eslinger PJ (2005a) The moral affiliations of disgust: a functional MRI study. Cogn Behav Neurol 18(1):68–78. doi:pii:00146965-200503000-00008

Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J (2005b) Opinion: the neural basis of human moral cognition. Nat Rev Neurosci 6(10):799–809. doi:10.1038/nrn1768

Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. Proc Natl Acad Sci USA 103(42):15623–15628. doi:10.1073/pnas.0604475103

Moran JM, Young LL, Saxe R, Lee SM, O'Young D, Mavros PL, Gabrieli JD (2011) Impaired theory of mind for moral judgment in high-functioning autism. Proc Natl Acad Sci USA 108(7):2688–2692. doi:10.1073/pnas.1011734108

Moretto G, Ladavas E, Mattioli F, di Pellegrino G (2010) A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. J Cogn Neurosci 22(8):1888–1899. doi:10.1162/jocn.2009.21367

Müller VI, Habel U, Derntl B, Schneider F, Zilles K, Turetsky BI, Eickhoff SB (2011) Incongruence effects in crossmodal emotional integration. Neuroimage 54(3):2257–2266. doi:10.1016/j.neuroimage.2010.10.047

Phelps EA, O'Connor KJ, Gatenby JC, Gore JC, Grillon C, Davis M (2001) Activation of the left amygdala to a cognitive representation of fear. Nat Neurosci 4(4):437–441. doi:10.1038/86110

Piaget J (1932) The moral judgment of the child. Kegan Paul, London

Pizarro DA, Bloom P (2003) The intelligence of the moral intuitions: comment on Haidt (2001). Psychol Rev 110(1):193–196

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10(2):59–63. doi:10.1016/j.tics.2005.12.004

Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind. Behav Brain Sci 1(4):515–526

Raine A, Yang Y (2006) Neural foundations to moral reasoning and antisocial behavior. Soc Cogn Affect Neurosci 1(3):203–213. doi:10.1093/scan/nsl033

Rosenthal R (1979) The file drawer problem and tolerance for null results. Psychol Bull 86(3):638–641

Sander D, Grafman J, Zalla T (2003) The human amygdala: an evolved system for relevance detection. Rev Neurosci 14(4):303–316

Saper CB (2000) Pain as a visceral sensation. Prog Brain Res 122:237–243

Saxe R, Kanwisher N (2003) People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". Neuroimage 19(4):1835–1842. pii:S1053811903002301

Schacter DL, Addis DR, Buckner RL (2007) Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci 8(9):657–661. doi:10.1038/nrn2213

Schilbach L (2010) A second-person approach to other minds. Nat Rev Neurosci 11(6):449. doi:10.1038/nrn2805-c1

Schilbach L, Eickhoff SB, Mojzisch A, Vogeley K (2008a) What's in a smile? Neural correlates of facial embodiment during social interaction. Soc Neurosci 3(1):37–50. doi:10.1080/17470910701563228

Schilbach L, Eickhoff SB, Rotarska-Jagiela A, Fink GR, Vogeley K (2008b) Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. Conscious Cogn 17(2):457–467. doi:10.1016/j.concog.2008.03.013

Schilbach L, Wilms M, Eickhoff SB, Romanzetti S, Tepest R, Bente G, Shah NJ, Fink GR, Vogeley K (2010) Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. J Cogn Neurosci 22(12):2702–2715. doi:10.1162/jocn.2009.21401

Schilbach L, Eickhoff SB, Cieslik EC, Kuzmanovic B, Vogeley K (2012) Shall we do this together? Social gaze influences action control in a comparison group, but not in individuals with high-functioning autism. Autism. doi:10.1177/1362361311409258

Shamay-Tsoory SG, Tibi-Elhanany Y, Aharon-Peretz J (2006) The ventromedial prefrontal cortex is involved in understanding affective but not cognitive theory of mind stories. Soc Neurosci 1(3–4):149–166. doi:10.1080/17470910600985589

Shamay-Tsoory SG, Aharon-Peretz J, Perry D (2009) Two systems for empathy: a double dissociation between emotional and

cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain 132:617–627. doi:10.1093/Brain/Awn279

Singer T, Lamm C (2009) The social neuroscience of empathy. Ann N Y Acad Sci 1156:81–96. doi:10.1111/j.1749-6632.2009.04418.x

Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD (2004) Empathy for pain involves the affective but not sensory components of pain. Science 303(5661):1157–1162. doi:10.1126/science.1093535

Soderstrom H (2003) Psychopathy as a disorder of empathy. Eur Child Adolesc Psychiatry 12:249–252

Spengler S, von Cramon DY, Brass M (2009) Control of shared representations relies on key processes involved in mental state attribution. Hum Brain Mapp 30(11):3704–3718. doi:10.1002/hbm.20800

Spreng RN, Mar RA, Kim AS (2009) The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. J Cogn Neurosci 21(3):489–510. doi:10.1162/jocn.2008.21029

Suddendorf T, Corballis MC (2007) The evolution of foresight: What is mental time travel, and is it unique to humans? Behav Brain Sci 30(3):299–313. doi:10.1017/S0140525X07001975 (discussion 313–351)

Tangney JP, Stuewig J, Mashek DJ (2007) Moral emotions and moral behavior. Annu Rev Psychol 58:345–372

Tomasello M (2001) The cultural origins of human cognition. Harvard University Press, Cambridge

Tulving E (1983) Elements of Episodic Memory. Clarendon Press, Oxford

Tulving E (1985) Memory and consciousness. Can Psychol 26:1–12

Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA (2002) Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage 16(3):765–780. doi:10.1006/nimg.2002.1131

Turkeltaub PE, Eickhoff SB, Laird AR, Fox M, Wiener M, Fox P (2011) Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. Hum Brain Mapp. doi:10.1002/hbm.21186

Van Overwalle F (2009) Social cognition and the brain: a meta-analysis. Hum Brain Mapp 30(3):829–858. doi:10.1002/hbm.20547

Van Overwalle F (2011) A dissociation between social mentalizing and general reasoning. Neuroimage 54(2):1589–1599. doi:10.1016/j.neuroimage.2010.09.043

Vogt BA (2005) Pain and emotion interactions in subregions of the cingulate gyrus. Nat Rev Neurosci 6(7):533–544. doi:10.1038/nrn1704

Wagner U, N'Diaye K, Ethofer T, Vuilleumier P (2012) Guilt-Specific Processing in the Prefrontal Cortex. Cereb Cortex. doi:10.1093/cercor/bhr016

Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G (2003) Both of us disgusted in My Insula: The common neural basis of seeing and feeling disgust. Neuron 40(3):655–664. doi:10.1016/S0896-6273(03)00679-2

Young L, Saxe R (2008) The neural basis of belief encoding and integration in moral judgment. Neuroimage 40(4):1912–1920. doi:10.1016/j.neuroimage.2008.01.057

Young L, Saxe R (2009) Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. Neuropsychologia 47(10):2065–2072. doi:10.1016/j.neuropsychologia.2009.03.020

Young L, Cushman F, Hauser M, Saxe R (2007) The neural basis of the interaction between theory of mind and moral judgment. Proc Natl Acad Sci USA 104(20):8235–8240. doi:10.1073/pnas.0701408104

Young L, Bechara A, Tranel D, Damasio H, Hauser M, Damasio A (2010a) Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. Neuron 65(6):845–851. doi:10.1016/j.neuron.2010.03.003

Young L, Camprodon JA, Hauser M, Pascual-Leone A, Saxe R (2010b) Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proc Natl Acad Sci USA 107(15):6753–6758. doi:10.1073/pnas.0914826107

Zahn-Waxler C, Radke-Yarrow M, Wagner E, chapman M (1992) Development of concern for others. Dev Psychol 28:126–136