Opinion piece

Author for correspondence:
C. Lamm
e-mail: claus.lamm@univie.ac.at

From shared to distinct self–other representations in empathy: evidence from neurotypical function and socio-cognitive disorders

C. Lamm[1], H. Bukowski[1] and G. Silani[2]

[1]Social, Cognitive and Affective Neuroscience Unit, Department of Basic Psychological Research and Research Methods, and [2]Department of Applied Psychology: Health, Development, Enhancement and Intervention, University of Vienna, Vienna, Austria

CL, 0000-0002-5422-0653; HB, 0000-0001-9412-1855; GS, 0000-0002-4284-3618

Neuroscientific research has identified two fundamental components of empathy: shared emotional representations between self and other, and self–other distinction. The concept of shared representations suggests that during empathy, we co-represent another person's affect by engaging brain and bodily functions underpinning the first-hand experience of the emotion we are empathizing with. This possible grounding of empathy in our own emotional experiences explains the necessity for self–other distinction, which is the capacity to correctly distinguish between our own affective representations and those related to the other. In spite of the importance of these two components in empathy, several aspects still remain controversial. This paper addresses some of them and focuses on (i) the distinction between shared activations versus representations, raising the question what shared representations entail in terms of the underlying neural mechanisms, (ii) the possible mechanisms behind self–other distinction in the cognitive and the affective domains, and whether they have distinct neural underpinnings and (iii) the consequences associated with a selective impairment of one of the two components, thereby addressing their importance in mental disorders such as autism spectrum disorders, psychopathy and alexithymia.

1. From shared activations to shared representations

The seminal discovery of Singer *et al.* [1] that empathy for pain results in brain activations overlapping with those engaged when being in pain oneself has triggered a persistent debate regarding what these overlapping or shared activations imply in terms of the underlying neural functions and psychological mechanisms. Initially, the majority of scholars have welcomed the fact that similar parts of the brain are active during first-hand and empathic experiences as support for claims grounded in simulation theory and embodied social cognition that we 'reactivate' or use our own emotion systems in order to sense and feel what another person is sensing and feeling.

A major drawback of neuroimaging methods is, however, that the relationship between neural activations and mental representations—and the cognitive and affective processes supposedly engaged in an experimental condition—is not always straightforward ([2], for in-depth discussion). In the light of the method's limitations, it is mainly the existence of shared activations ('the language of the brain') but not necessarily of shared representations ('the language of the mind/cognition'), that could be derived by previous social neuroscience research on empathy. For instance, the repeatedly reported activation overlap during first-hand pain and pain empathy in the anterior midcingulate cortex (aMCC) and bilateral anterior insular cortex (aIns; [2], for review) has consistently been linked to the affective-motivational component of first-hand pain (e.g. [3]). It has thus been suggested that empathy involves the sharing of pain affect, and that understanding others'

emotions indeed may be based on an embodied simulation of other people's emotions grounded in one's own emotion experiences. Apart from the inherent methodological limitations of neuroimaging, there are several empirical challenges for such claims.

For instance, patients suffering from congenital insensitivity to pain who might not activate aMCC and aIns when being in pain themselves (though this was actually never shown) also activate aMCC and aIns when seeing others suffering somatic pain [4]. Moreover, several studies showed activation in aMCC and aIns [5,6] when observers watched medical procedures such as pinpricks which only appeared aversive, but in reality were not painful for the depicted person (e.g. because he/she suffered from a rare neurological disorder or his/her hand had been anaesthetized). This seems difficult to reconcile with the interpretation that these activations represent a sharing of the other person's actual feelings, because in this case there were actually no feelings to be shared.

Findings such as these mainly raise questions about the *specificity* of shared activations, and are in line with emerging concepts that aMCC and in particular aIns might be domain-general hubs related to saliency detection and emotional awareness (e.g. [7]). Hence, rather than representing the other's affect, activation of these areas might stem from the phenomenon that seeing someone in a painful situation is a strong and salient social cue arousing the observer, and preparing him or her to respond to a potential threat. Moreover, the aversiveness of the observed situation may trigger self-related or 'personal distress' signals, which in the social psychology literature have long been recognized as one of the key components of responses to the emotions of others ([8], for review). This would explain activations during medical procedures in the absence of actual pain as a lack of adequate or timely regulation of one's own aversive response.

Moreover, the *sensitivity* of shared activations has recently been questioned in the domain of social rejection which also heavily draws on the concept of shared representations [9]. Using multivariate pattern analyses (MVPAs) of fMRI data, it has been shown that shared activations between somatic pain and the 'pain' of social rejection stem from distinct underlying activation patterns [10], and similarly negative findings for pain empathy seem to await confirmation (Tor Wager 2015, personal communication; but see also [11]). MVPA certainly raises important questions regarding the validity of traditional mass univariate analyses targeting shared representations, and it holds great promise for advancing our understanding of whether shared task activations indeed imply shared representations. However, MVPA suffers from the same limitations as mass univariate fMRI in terms of how accurately they can depict the underlying neural computations [12].

This general limitation should also remind us that the reliance on any single method is ill-advised. While fMRI certainly allows important new insights into the neural bases of social cognition, real progress seems to require a combination of methods and the generation of converging evidence— i.e. different methods or paradigms advocating the same type of conclusions. This includes, foremost, the use of methods allowing causal rather than just correlational inferences— such as for instance those used in lesion, brain stimulation or psychopharmacological studies. Regarding lesion studies, there is accumulating evidence that damage to areas such as aMCC and aIns results in deficits in empathy [13]. However, focal lesions of aMCC and aIns are not only scarce, but their specificity is also limited as pervasive damage to these areas might simply affect domain-general processes such as salience detection or emotional awareness. 'Virtual lesions' by means of brain stimulation are difficult to attain for aMCC and aIns, owing to their location far away from the cortical surface, where the brain stimulation coil or electrode is placed. In the light of these limitations, we have recently proposed another research approach which attempts to generate more mechanistic insights into the neural bases of empathy and of shared activations. This approach relies on the combination of causal experimental and psychopharmacological manipulations, and uses the induction of placebo analgesia to test whether the associated reduction in first-hand pain also results in a reduction of pain empathy. Based on self-reported pain and empathy ratings, this was clearly and reliably the case in three independent studies and with three different neuroscience methods producing converging evidence. More specifically, placebo analgesia reduced a pain-related event-related potential (ERP) component during empathy for pain that was likely to originate from aMCC [14], and it also reduced fMRI activation in aMCC and aIns during both first-hand and empathy for pain [15]. Notably, using the opioid antagonist naltrexone to pharmacologically block the effects of placebo analgesia also blocked the effects of placebo analgesia on empathy for pain. Taken together, these studies indicate through causal experimental and psychopharmacological manipulations that empathy for pain is grounded in the first-hand experience of pain. They also indirectly indicate that the opioid system, which plays a central role in pain regulation, may also be engaged in empathy for pain. This however is an assumption that needs more specific testing. Moreover, these findings are as yet not fully conclusive as regards which type of affective representations—specifically those related to pain, or just negative affect—are modulated by placebo analgesia, as the latter has also been shown to affect aversive emotions in general. However, we propose that approaches in which first-hand emotion experiences are experimentally manipulated to measure how this affects empathic experiences hold great potential for a more definite understanding of shared representations in empathy research (see also [16], for a similar idea).

In summary, despite more than a decade of intense scrutiny, definite conclusions into what shared activations during empathy really imply in terms of the underlying neuro-cognitive and -affective processes are still lacking. While there are some interpretations that they represent rather unspecific domain-general processes (related to salience, or personal distress), more recent causal approaches seem to advocate a more specific role in coding actual affective representations. What is not contested is that empathy entails a robust affective response, which is expressed not only at the neural but also at the bodily level. This response might be constituted of a mix of self- and other-related processes, which highlights the importance of self–other distinction, to which we turn now.

2. Self–other distinction and empathy

Self–other distinction is referred to as a key mechanism in many fields of social cognition, and generally describes the ability to distinguish between the representations of our own actions, perceptions, sensations and emotions, and those of others. In empathy, accurate self–other distinction is important either to avoid the occurrence of personal distress or to prevent

our own affective state egocentrically biasing how we empathize with others. While the former aspect has been intensely investigated ([8], for review), the phenomenon of emotional egocentricity bias has only recently been introduced in the scientific literature. Using a novel experimental paradigm in which two persons simultaneously underwent an affective touch of incongruent valence, it was shown that the valence of one's own affective touch egocentrically biased empathic judgements of the affect touch of the other person [17]. While this bias was relatively small in healthy young adults, temporarily inhibiting the right supramarginal gyrus (rSMG) substantially increased it, and so did acute stress in a male sample [18]. This suggests that under normal conditions, affective self–other distinction enables us to overcome egocentric interference with empathy, while impeding a brain region implementing it or reduced cognitive capacity due to acute stress leads to substantial empathic misattributions. Interestingly, rSMG is located anterior to the right temporo-parietal junction (rTPJ), a brain area that has classically been associated with self–other distinction in cognitive and motor domains, but also during empathy (e.g. [19,20]). This raises the question as to whether self–other distinction in different domains is underpinned by a unitary or distinct mechanisms.

In terms of the neural correlates, overcoming emotional egocentricity in the affective domain seems consistently and specifically underpinned by rSMG [21]. Moreover, connectivity analyses have shown that rSMG and rTPJ show distinct connectivity profiles, with the former having stronger connections to 'emotion areas' such as the aMCC and aIns [22]. While suggestive, this however does not necessarily imply different computational mechanisms for 'cognitive' and 'affective' self–other distinction. Rather, it might be that the two neighbouring regions serve the same function, but are differentially engaged because of the different (affective versus cognitive) content they access and regulate. Likewise, the engagement of rSMG might relate to the specific task constraints of overcoming emotional egocentricity when being in an emotional state oneself—as indicated by studies in which empathizers in a neutral state consistently activated more posterior parts in rTPJ 'proper', and not in rSMG ([23,24], for recent meta-analyses). Generally, the recruitment of rTPJ across a range of social-cognitive tasks tapping into self–other distinction renders this area—notwithstanding possible sub-divisions—a major target for gaining a better understanding of self–other distinction and its neuro-cognitive mechanisms [20,25].

Notably, rTPJ engagement is far from restricted to the domain of social cognition. Thanks to its strong connectivity with prefrontal and posterior parietal structures [22,25], many functions are associated with this area and three of them seem of particular relevance for the present discussion: rTPJ has been associated with attentional reorienting to salient or task-relevant events [26], it may be engaged in the comparison of internal expectations with external events [20,27], and it seems to allow a 'rich', multimodal representation of the social context [25]. We propose that the combination of these three functions is what underpins self–other distinction. More specifically, the rTPJ has been described as a nexus connecting interoceptive and exteroceptive information [25]. This nexus would, however, be overwhelmed if not equipped with mechanisms to detect, select, and keep track of cognitively or behaviourally relevant events. These mechanisms enable the rTPJ to represent the social context in a way in which both

self and other functions as two independent agents, and to make predictions about how events will evolve and how agents will behave. These predictions and the resulting internal models direct our attention towards behaviourally relevant social stimuli, and in particular to those indicating a violation of our models. The resulting acquisition of novel information is used to adjust self- and other-related representations, and/or to update the internal models. For example, to avoid bumping into someone when crossing a place crowded with people, we need to be able to detect, select and keep track of the relevant agents, and to predict their moves into different directions. This, among others, can be derived from the direction of their gaze and when and whether they also gaze at us. However, a single gaze may signal many meanings, and contextual integration is critical for correct understanding. Hence, it is not surprising that online gaze processing also consistently engages TPJ (and mPFC; [28,29], for reviews). Such findings are in line with a model of TPJ function which suggests that this brain area supports the integration of the vast array of intero- and exteroception information required by seemingly simple social interactions and situations.

In the emotional domain, rSMG seems to fulfil a similar integration function, which might be based on a similar mechanism as suggested by the research on affective touch referred to above [17]. Notably, in the experimental paradigm used, empathy had to be derived from a picture depicting the (pleasant or unpleasant) object the other person was being touched with, while participants underwent touch on their right hand and concurrently also saw a picture of the object they were being touched with. When assessing effective connectivity of the rSMG during incongruent trials (i.e. in situations where self–other distinction demands were highest), connectivity was increased with bilateral higher order visual and left primary somatosensory cortex, i.e. contralateral to the touched hand. The increased coupling of rSMG with these two brain areas could therefore be interpreted as rSMG providing access to self- and other-related representations, and allowing to switch between or to integrate them. Extending these findings [30], a lower emotional egocentricity bias in adults compared with children was associated with higher connectivity between the rSMG and the left dlPFC. However, this was only the case when empathic judgements were explicitly required by means of empathy ratings. The increased coupling of rSMG with a cortical control structure might therefore reflect the higher demand to 'overrule' the interference from one's own emotional state when requiring an explicit behavioural response.

Irrespective of whether self–other distinction relies on a unitary mechanism, what is missing is an elaborated model of how self–other distinction is actually achieved in cases of mismatching self- and other-related representations. Several options are conceivable, but none of them has as yet been explicitly addressed by previous research. One option is to amplify either representations related to the self or to the other, depending on task requirements. For instance, in automatic imitation tasks where one has to execute a movement and simultaneously experiences interference from another person's mismatching movement [31], rTPJ and prefrontal structures may support each other to reinforce the self-related action goal over the intrusion of the other person's goal. When overcoming emotional egocentricity, in which the self intrudes on the representation of the other, the same mechanism might be at play, but in this case it amplifies representations related to the other rather than the self.

Again, this might be achieved by a combination of temporo-parietal and prefrontal control structures. Another option is that self–other distinction is achieved not by amplifying those representations which are crucial for a task, but by suppressing those which interfere with it. Finally, self–other distinction may be achieved neither by amplifying nor suppressing self- or other-related representations, but by a more accurate 'tagging' of which representations belong to the self and which to the other. This strategy would therefore not require modulating the existing representations, but simply require 'keeping them more clear' and allowing a more efficient switching between them.

In summary, self–other distinction is a critical mechanism of empathy, but a detailed understanding of its neuro-cognitive and -affective mechanisms is still missing. The integrative 'nexus' model [25] on which we have expanded here is certainly a useful heuristic, but many questions regarding the function of TPJ and of self–other distinction have currently only received crude answers, including the actual computations by which it is achieved. In this respect, it will be interesting to see how the propositions for a 'second-person neuroscience' and the use of more interactive and ecological experimental paradigms [32] will increase our understanding of self–other distinction and TPJ function. Based on the model proposed, we predict that the use of 'online' interactive tasks will increase the amount of contextual and of self- and other-related information that needs to be processed in the next generation of social cognition paradigms. This should engage the integrative processes pertaining to the TPJ to a larger extent, and it will certainly provide added opportunities to investigate failures of self–other distinction and the dynamic updating of shared representations.

3. Shared representations and self–other distinction in socio-cognitive disorders

Research on clinical populations provides a fundamental source of knowledge about the mechanisms behind complex cognitive functions and behaviours. Empathy is a clear example where clinical populations help us to understand the consequences associated with the alteration/dysfunctioning of this social emotion. Autism spectrum disorders (ASD) and psychopathy represent the most interesting clinical examples, because they have been considered to carry very different empathic deficits.

ASD are characterized by a developmental disorder which entails altered social cognition and communication and restricted, repetitive and stereotyped behaviours [33]. Persons with ASD are often described as lacking empathy. For instance, in questionnaire studies they usually obtain low scores on the empathizing axis and high scores on the systematizing axis [34], and report reduced empathic concern [35]. Recently, however, the assertion of a global deficit of empathy observed in ASD has been questioned when assessing more fine-grained distinctions of the different components of empathy, suggesting rather preserved shared representations [36]. A further important differentiation in the characterization of the empathic deficits observed in autism is its high comorbidity with alexithymia. Alexithymia has been described as a subclinical phenomenon marked by difficulties in identifying and describing feelings, as well as in distinguishing them from bodily sensations of emotional arousal [37]. Although not included in the formal diagnosis of autism, recent studies

have demonstrated severe degrees of comorbidity affecting up to 50% of persons with ASD (e.g. [38]). Recently, for example, it has been shown that the inability of ASD participants to understand and identify their emotions (determined by the self-reported level of alexithymia) is associated with a reduced response in empathizing with the pain of others, and this was mirrored on the neural level by hypoactivation of the aIns [39]. These findings have two important implications: first, they suggest that a correct representation of our own emotions may be a necessary condition for experiencing shared emotion representations, and second, that it is not ASD *per se*, but alexithymia that lies at the core of impaired emotion sharing.

Therefore, if it is not ASD specifically which affects emotion sharing, the question remains which deficits in empathy are specifically related to ASD. In addition to the repeatedly reported deficits in Theory of Mind (e.g. [40]), and consequently the ability (or lack of) to correctly adopt the perspective of other people, it is an interesting hypothesis that what may be impaired in ASD is the ability to correctly attribute emotional responses to the other person—i.e. the ability not to confuse self- and other-related emotional responses. Impairments in self–other distinction would predict greater emotional contagion and increased personal distress. Interestingly, there is pervasive evidence of greater personal distress in participants with ASD when witnessing others' negative emotions (see [41], for review). As suggested in the previous section, being limited in forming an accurate representation of both the self and the other as two independent agents, and to switch our attention from one to the other will result in a failure to update our internal model and to integrate or downregulate our own emotional response. This could possibly lead to inadequate social behaviours, motivated by reducing one's distress rather than attending to that of the other [8]. In this context, it is worth noting that higher levels of self-reported personal distress when witnessing others' pain have been reported as a key feature of alexithymia as well [42–44]. This leaves the possibility that it is the comorbidity with alexithymia that accounts for the level of personal distress observed in the ASD population.

Further initial evidence for self–other distinction deficits in autism is also found in the action domain, as ASD participants have been reported to show hyperimitation in an automatic imitation task considered to tap into self–other distinction [45,46], and also show reduced rTPJ activation in this task [45]. Moreover, it is noteworthy that the social-cognitive deficits reported in ASD have been argued to originate from impaired domain-general attentional processes associated with the TPJ [47]. This would be in line with the finding that TPJ has repeatedly been reported as less active in ASD participants compared with healthy controls [48,49]. Because TPJ as outlined above plays a key role in self–other distinction, this also provides further indication that deficits in self–other distinction may be at the core of the empathy deficits observed in ASD. On the other hand, given the recent evidence that rSMG rather than rTPJ may be specifically involved in affective self–other distinction, one would rather expect changes in rSMG. There is no evidence for this at the moment, which is not surprising, considering the relatively recent introduction of experimental paradigms tailored to measure self–other distinction in the affective domain [17].

Future studies are therefore needed to more specifically address the question whether within ASD, with or without comorbid alexithymia, self–other distinction during empathic

responses is altered. These studies should use paradigms tailored to investigate emotional versus cognitive-motor aspects of self–other distinction in order to be able to distinguish rSMG from rTPJ functions and self–other distinction related to affective versus cognitive content. Moreover, they should assess comorbidity with levels of alexithymia both in ASD and neurotypical populations, and they should also vary the level of personal distress to investigate how becoming overwhelmed by the other's negative emotions aggravates self–other distinction deficits in ASD.

Psychopathy is another disorder characterized by severe impairments of empathy. According to one prevalent conceptualization [50], individuals scoring high on the affective and interpersonal factor of psychopathy are characterized by empathy deficits, shallow affect and lack of remorse or guilt. Unlike in ASD, the empathy deficit observed in psychopathy has been associated with altered shared emotion representations [51].

Over the past decades, a handful of neuroimaging studies have started to investigate how psychopaths respond to the emotions of others. This has revealed a general inability to affectively and empathetically respond to others' emotions, as indicated by reduced physiological (SCR, heart rate) and neural activation in areas such as the insula and the amygdala (e.g. [52,53]). It has thus been suggested that the altered ability to experience negative emotions lies at the core of the impairments in correctly recognizing and responding to others' emotions in psychopathy [54,55]. The extent of this impairment is far from being understood, though. It has, for instance, been observed that participants with high psychopathy were able to show a typical response within the network involved in empathy for pain when asked to take a 'self-perspective' (i.e. a perspective requiring them to imagine explicitly to be in the place of a person suffering pain [56]). Conversely, when taking a more detached 'other-perspective', psychopaths exhibited an atypical pattern of brain activation. This suggests a dissociation between self- and other-related emotional experiences. Somewhat contradictorily, though, it has also been shown that in spite of a blunted automatic emotional response to others' emotions (including positive ones), psychopaths display 'normal' empathy-related brain responses when explicitly instructed to empathize with others [57]. This opens the possibility that top-down mechanisms can compensate for the absence of the automatic empathic response, by recruiting brain regions coding for shared affect.

Moreover, psychopaths have been described as having an extraordinary ability to manipulate and charm others to achieve their goals [50], suggesting a superior capacity to represent their own and other-related states. In support of this hypothesis, research in the last decades has shown that psychopaths are proficient in cognitively representing both others' desires, beliefs and intentions (as documented by their normal capacities—and related brain activations—to pass Theory of Mind tasks, also in the affective domain [58]), and their own emotional experiences (as indicated by the absence of alexithymic traits among psychopaths with deficits in the affective-interpersonal domain [59]). Interestingly, a recent forensic study suggested

that psychopathic inmates seemed to possess advanced skills to predict what kind of self-reported empathic responses to others' pain were expected from them [60]. More specifically, inmates with high psychopathic traits reported cognitive-evaluative responses to the pain of others that matched those of normal controls, while inmates with low traits failed to do so. Both however showed similarly blunted autonomic responses to others' pain. The intact and maybe even superior ability to represent others' mental states may be a possible reason for the ability to correctly judge others' emotions in spite of reduced affective responding. For instance, Book et al. [61] observed individuals with higher psychopathic traits to make more accurate judgements of emotional intensity and vulnerability than non-psychopaths by relying on socially relevant cues such as body language and contextual cues.

Finally, similarly to ASD, studies specifically targeting self–other distinction during empathy and in other domains in psychopathy are missing. It is therefore not clear which aspect of self–other distinction psychopaths are particularly skilled at. Apart from possibly superior abilities for perspective taking, they (also) might be very skilled at distancing themselves from others. Future studies must address this issue, by investigating for instance how psychopaths perform in emotional egocentricity paradigms requiring the suppression of preponderant self-related representations in order to correctly empathize with others (e.g. [17]), or when the inhibition of other-related representations is necessary in order to correctly perform a task (e.g. [45]).

4. General conclusion

The aim of this paper was to provide an overview of two key components of empathy, and to discuss how they might be implemented in neurotypical individuals and affected in socio-cognitive disorders. In spite of the repeatedly indicated lack of more definite answers to questions such as how the mechanisms of affective sharing and self–other distinction are exactly achieved, the following findings can be derived from existing evidence: (i) empathy, as defined here, is a complex and multi-faceted social emotion; (ii) shared representations between self and others' emotions and the correct distinction between these two seem to be necessary conditions for empathy; and (iii) a deficit in self–other distinction could lead to higher levels of personal distress and reduced other-oriented responses (such as in ASD), or a deficit in affective sharing could lead to callous and unemotional behaviours (such as in psychopathy). We expect that future research attaining a more fine-grained delineation of the exact neurocomputational mechanisms will also inform more specifically about the roots of the empathy deficits in clinical disorders, thus opening up more tailored options for intervention and treatment.

References

1. Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD. 2004 Empathy for pain involves the affective but not the sensory components of pain. Science 303, 1157–1161. (doi:10.1126/science.1093535)

2. Lamm C, Majdandzic J. 2015 The role of shared neural activations, mirror neurons, and morality in

empathy – a critical comment. *Neurosci. Res.* **90C**, 15–24. (doi:10.1016/j.neures.2014.10.008)

3. Tracey I, Mantyh PW. 2007 The cerebral signature for pain perception and its modulation. *Neuron* **55**, 377–391. (doi:10.1016/j.neuron.2007.07.012)

4. 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**, 203–212. (doi:10.1016/j.neuron.2008.11.023)

5. Lamm C, Meltzoff AN, Decety J. 2010 How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *J. Cogn. Neurosci.* **22**, 362–376. (doi:10.1162/jocn.2009.21186)

6. Lamm C, Nusbaum HC, Meltzoff AN, Decety J. 2007 What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *PLoS ONE* **12**, e1292. (doi:10.1371/journal.pone.0001292)

7. Legrain V, Iannetti GD, Plaghki L, Mouraux A. 2011 The pain matrix reloaded: a salience detection system for the body. *Prog. Neurobiol.* **93**, 111–124. (doi:10.1016/j.pneurobio.2010.10.005)

8. Decety J, Lamm C. 2011 Empathy versus personal distress: recent evidence from social neuroscience. In *The social neuroscience of empathy* (eds J Decety, W Ickes), pp. 199–214. Cambridge, MA: MIT Press.

9. Eisenberger NI. 2015 Social pain and the brain: controversies, questions, and where to go from here. *Annu. Rev. Psychol.* **66**, 601–629. (doi:10.1146/annurev-psych-010213-115146)

10. Woo CW, Koban L, Kross E, Lindquist MA, Banich MT, Ruzic L, Andrews-Hanna JR, Wager TD. 2014 Separate neural representations for physical pain and social rejection. *Nat. Commun.* **5**, 5380. (doi:10.1038/ncomms6380)

11. Corradi-Dell'Acqua C, Hofstetter C, Vuilleumier P. 2011 Felt and seen pain evoke the same local patterns of cortical activity in insular and cingulate cortex. *J. Neurosci.* **31**, 17 996–18 006. (doi:10.1523/JNEUROSCI.2686-11.2011)

12. Dubois J, de Berker AO, Tsao DY. 2015 Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J. Neurosci.* **35**, 2791–2802. (doi:10.1523/JNEUROSCI.4037-14.2015)

13. Hillis AE. 2014 Inability to empathize: brain lesions that disrupt sharing and understanding another's emotions. *Brain* **137**, 981–997. (doi:10.1093/brain/awt317)

14. Rütgen M, Seidel EM, Riecansky I, Lamm C. 2015 Reduction of empathy for pain by placebo analgesia suggests functional equivalence of empathy and first-hand emotion experience. *J. Neurosci.* **35**, 8938–8947. (doi:10.1523/JNEUROSCI.3936-14.2015)

15. Rütgen M, Seidel E-M, Silani G, Riečanský I, Hummer A, Windischberger C, Petrovic P, Lamm C. 2015 Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proc. Natl Acad. Sci. USA* **112**, E5638–E5646. (doi:10.1073/pnas.1511269112)

16. Olsson A, McMahon K, Papenberg G, Zaki J, Bolger N, Ochsner KN. In press. Vicarious fear learning depends on empathic appraisals and trait empathy. *Psychol. Sci.*

17. Silani G, Lamm C, Ruff CC, Singer T. 2013 Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *J. Neurosci.* **33**, 15 466–15 476. (doi:10.1523/JNEUROSCI.1488-13.2013)

18. Tomova L, von Dawans B, Heinrichs M, Silani G, Lamm C. 2014 Is stress affecting our ability to tune into others? Evidence for gender differences in the effects of stress on self–other distinction. *Psychoneuroendocrinology* **43**, 95–104. (doi:10.1016/j.psyneuen.2014.02.006)

19. Santiesteban I, Banissy MJ, Catmur C, Bird G. 2012 Enhancing social ability by stimulating right temporoparietal junction. *Curr. Biol.* **22**, 2274–2277. (doi:10.1016/j.cub.2012.10.018)

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

21. Steinbeis N. 2016 The role of self–other distinction in understanding others' mental and emotional states: neurocognitive mechanisms in children and adults. *Phil. Trans. R. Soc. B* **371**, 20150074. (doi:10.1098/rstb.2015.0074)

22. Mars RB, Sallet J, Schuffelgen U, Jbabdi S, Toni I, Rushworth MFS. 2012 Connectivity-based subdivisions of the human right 'temporoparietal junction area': evidence for different areas participating in different cortical networks. *Cereb. Cortex* **22**, 1894–1903. (doi:10.1093/cercor/bhr268)

23. Bzdok D, Schilbach L, Vogeley K, Schneider K, Laird AR, Langner R, Eickhoff SB. 2012 Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct. Funct.* **217**, 783–796. (doi:10.1007/s00429-012-0380-y)

24. 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**, 2492–2502. (doi:10.1016/j.neuroimage.2010.10.014)

25. Carter RM, Huettel SA. 2013 A nexus model of the temporal–parietal junction. *Trends Cogn. Sci.* **17**, 328–336. (doi:10.1016/j.tics.2013.05.007)

26. Corbetta M, Patel G, Shulman GL. 2008 The reorienting system of the human brain: from environment to theory of mind. *Neuron* **58**, 306–324. (doi:10.1016/j.neuron.2008.04.017)

27. Geng JJ, Vossel S. 2013 Re-evaluating the role of TPJ in attentional control: contextual updating? *Neurosci. Biobehav. Rev* **37**, 2608–2620. (doi:10.1016/j.neubiorev.2013.08.010)

28. Hamilton AF de C. 2016 Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Phil. Trans. R. Soc. B* **371**, 20150080. (doi:10.1098/rstb.2015.0080)

29. Pfeiffer UJ, Vogeley K, Schilbach L. 2013 From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev* **37**, 2516–2528. (doi:10.1016/j.neubiorev.2013.07.017)

30. Steinbeis N, Bernhardt BC, Singer T. 2015 Age-related differences in function and structure of rSMG and reduced functional connectivity with DLPFC explains heightened emotional egocentricity bias in childhood. *Soc. Cogn. Affect Neurosci.* **10**, 302–310. (doi:10.1093/scan/nsu057)

31. Brass M, Ruby P, Spengler S. 2009 Inhibition of imitative behaviour and social cognition. *Phil. Trans. R. Soc. B* **364**, 2359–2367. (doi:10.1098/rstb.2009.0066)

32. Schilbach L. 2016 Towards a second-person neuropsychiatry. *Phil. Trans. R. Soc. B* **371**, 20150081. (doi:10.1098/rstb.2015.0081)

33. Frith U, Happe F. 2005 Autism spectrum disorder. *Curr. Biol.* **15**, R786–R790. (doi:10.1016/j.cub.2005.09.033)

34. Baron-Cohen S, Wheelwright S. 2004 The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* **34**, 163–175. (doi:10.1023/B:JADD.0000022607.19833.00)

35. Lombardo MV, Barnes JL, Wheelwright SJ, Baron-Cohen S. 2007 Self-referential cognition and empathy in autism. *PLoS ONE* **2**, e883. (doi:10.1371/journal.pone.0000883)

36. Dziobek I, Rogers K, Fleck S, Bahnemann M, Heekeren HR, Wolf OT, Convit A. 2008 Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the Multifaceted Empathy Test (MET). *J. Autism Dev. Disord.* **38**, 464–473. (doi:10.1007/s10803-007-0486-x)

37. Nemiah JC, Freyberger H, Sifneos P. 1976 Alexithymia: a view of the psychosomatic process. *Mod. Trends Psychosom. Med.* **3**, 430–439.

38. Hill E, Berthoz S, Frith U. 2004 Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *J. Autism Dev. Disord.* **34**, 229–235. (doi:10.1023/B:JADD.0000022613.41399.14)

39. Bird G, Silani G, Brindley R, White S, Frith U, Singer T. 2010 Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain* **133**, 1515–1525. (doi:10.1093/brain/awq060)

40. Misra V. 2014 The social brain network and autism. *Ann. Neurosci.* **21**, 69–73. (doi:10.5214/ans.0972.7531.210208)

41. Smith A. 2009 The empathy imbalance hypothesis of autism: a theoretical approach to cognitive and emotional empathy in autistic development. *Psychol. Rec.* **59**, 489–510.

42. Grynberg D, Luminet O, Corneille O, Grèzes J, Berthoz S. 2010 Alexithymia in the interpersonal domain: a general deficit of empathy? *Pers. Individ. Differ.* **49**, 845–850. (doi:10.1016/j.paid.2010.07.013)

43. Moriguchi Y, Decety J, Ohnishi T, Maeda M, Mori T, Nemoto K, Matsuda H, Komaki G. 2007 Empathy and judging other's pain: an fMRI study of alexithymia. *Cereb. Cortex* **17**, 2223–2234. (doi:10.1093/cercor/bhl130)

44. Guttman H, Laporte L. 2002 Alexithymia, empathy, and psychological symptoms in a family context. *Compr. Psychiatry* **43**, 448–455. (doi:10.1053/comp.2002.35905)

45. Spengler S, Bird G, Brass M. 2010 Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biol. Psychiatry* **68**, 1148–1155. (doi:10.1016/j.biopsych.2010.09.017)

46. Sowden S, Koehne S, Catmur C, Dziobek I, Bird G. In press. Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: challenging the broken mirror theory. *Autism Res*. (doi:10.1002/aur.1511)

47. Leekam S. 2016 Social cognitive impairment and autism: what are we trying to explain? *Phil. Trans. R. Soc. B* **371**, 20150082. (doi:10.1098/rstb.2015.0082)

48. Fan YT, Chen C, Chen SC, Decety J, Cheng Y. 2014 Empathic arousal and social understanding in individuals with autism: evidence from fMRI and ERP measurements. *Soc. Cogn. Affect Neurosci.* **9**, 1203–1213. (doi:10.1093/scan/nst101)

49. Nummenmaa L, Engell AD, Von Dem Hagen E, Henson RN, Calder AJ. 2012 Autism spectrum traits predict the neural response to eye gaze in typical individuals. *Neuroimage* **59**, 3356–3363. (doi:10.1016/j.neuroimage.2011.10.075)

50. Hare RD, Neumann CS. 2005 Structural models of psychopathy. *Curr. Psychiatry Rep.* **7**, 57–64. (doi:10.1007/s11920-005-0026-3)

51. Blair R. 2007 Dysfunctions of medial and lateral orbitofrontal cortex in psychopathy. *Ann. NY Acad. Sci.* **1121**, 461–479. (doi:10.1196/annals.1401.017)

52. Blair RJ, Jones L, Clark F, Smith M. 1997 The psychopathic individual: a lack of responsiveness to distress cues? *Psychophysiology* **34**, 192–198. (doi:10.1111/j.1469-8986.1997.tb02131.x)

53. Marsh AA et al. 2008 Reduced amygdala response to fearful expressions in children and adolescents with callous-unemotional traits and disruptive behavior disorders. *Am. J. Psychiatry* **165**, 712–720. (doi:10.1176/appi.ajp.2007.07071145)

54. Brook M, Brieman CL, Kosson DS. 2013 Emotion processing in Psychopathy Checklist-assessed psychopathy: a review of the literature. *Clin. Psychol. Rev.* **33**, 979–995. (doi:10.1016/j.cpr.2013.07.008)

55. Bird G, Viding E. 2014 The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neurosci. Biobehav. Rev.* **47**, 520–532. (doi:10.1016/j.neubiorev.2014.09.021)

56. Decety J, Chen C, Harenski C, Kiehl KA. 2013 An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Front. Hum. Neurosci.* **7**, 489. (doi:10.3389/fnhum.2013.00489)

57. Meffert H, Gazzola V, den Boer JA, Bartels AA, Keysers C. 2013 Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. *Brain* **136**, 2550–2562. (doi:10.1093/brain/awt190)

58. Dolan M, Fullam R. 2004 Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychol. Med.* **34**, 1093–1102. (doi:10.1017/S0033291704002028)

59. Kroner DG, Forth AE. 1995 The Toronto alexithymia scale with incarcerated offenders. *Pers. Individ. Differ.* **19**, 625–634. (doi:10.1016/0191-8869(95)00116-N)

60. Pfabigan DM, Seidel EM, Wucherer AM, Keckeis K, Derntl B, Lamm C. 2015 Affective empathy differs in male violent offenders with high- and low-trait psychopathy. *J. Pers. Disord.* **29**, 42–61. (doi:10.1521/pedi_2014_28_145)

61. Book AS, Quinsey VL, Langford D. 2007 Psychopathy and the perception of affect and vulnerability. *Crim. Justice Behav.* **34**, 531–542. (doi:10.1177/0093854806293554)