OXFORD

UNDERSTANDING OTHER MINDS

PERSPECTIVES FROM DEVELOPMENTAL SOCIAL NEUROSCIENCE

EDITED BY SIMON BARON-COHEN,
HELEN TAGER-FLUSBERG AND
MICHAEL V. LOMBARDO

Chapter 12

Empathy and the brain

Cade McCall and Tania Singer

Feeling what others feel is basic to human social life. We wince when we see someone's finger sliced by a razor, when we see that person's face twist in pain, or even when we read about the event. Both physical cues and our imaginations are enough for us to infer and experience the affective states of others. These abilities have clear functional benefits, allowing us to learn from others' pain and to offer help and support when they're needed. While empathy is closely related to mentalizing about others' thoughts and intentions, sharing feelings is distinct from reading minds. In recent years, social neuroscience has made major strides in understanding empathy. Research on its neural representations and modulation has produced a complex picture. There is no single brain region underlying empathy, but a variety of networks that work together to produce (and prevent) vicarious feeling. Significant questions remain, particularly regarding the different domains of empathic experience, its developmental trajectories, and the translation of shared feelings into behavior. This chapter provides an overview of this work and highlights possible new directions for research.

Defining empathy

Definitions of empathy vary widely in their focus and breadth. Based heavily on groundbreaking work in psychology (Batson, 2009b; Eisenberg & Fabes, 1990; Wispe, 1986), social neuroscientists have honed in on a relatively specific construct for the purposes of research. One definition of empathy recently proposed by neuroscientists, for example, has four key components (de Vignemont & Singer, 2006; Decety & Jackson, 2004; Singer & Lamm, 2009). First, empathy refers to an affective state. Secondly, that state is elicited by the inference or imagination of another person's state. Thirdly, that state is isomorphic with the other person's state. Fourthly, the empathizer knows that the other person is the source of the state. In other words, empathy is the experience of vicariously feeling what another person is feeling without confounding the feeling with one's own direct experience (see Figure 12.1).

This definition distinguishes empathy from related phenomena. While mentalizing or cognitive perspective-taking may help us infer another person's affective state, it does not necessarily produce an affective state in ourselves. For example, mentalizing might produce the inference, "I see him smiling so he must be happy," while empathizing would produce the experience, "I am happy because he's happy." In gross terms, mentalizing represents more "cold" cognitive analysis of the scene and empathy the "warm" experiential response. Nevertheless, while these two constructs may be distinct on paper, mentalizing and empathy are closely related in mental life (Jackson, Brunet, Meltzoff, & Decety, 2006). For example, mentalizing plays a key part in providing the cues necessary to trigger empathic reactions. Conversely, empathic experience likely contributes to our mentalizing abilities by teaching us the meanings of specific affective cues.

Emotional contagion is another closely related phenomenon. In emotional contagion one "catches" the affective state of another person, but without awareness of the state's source

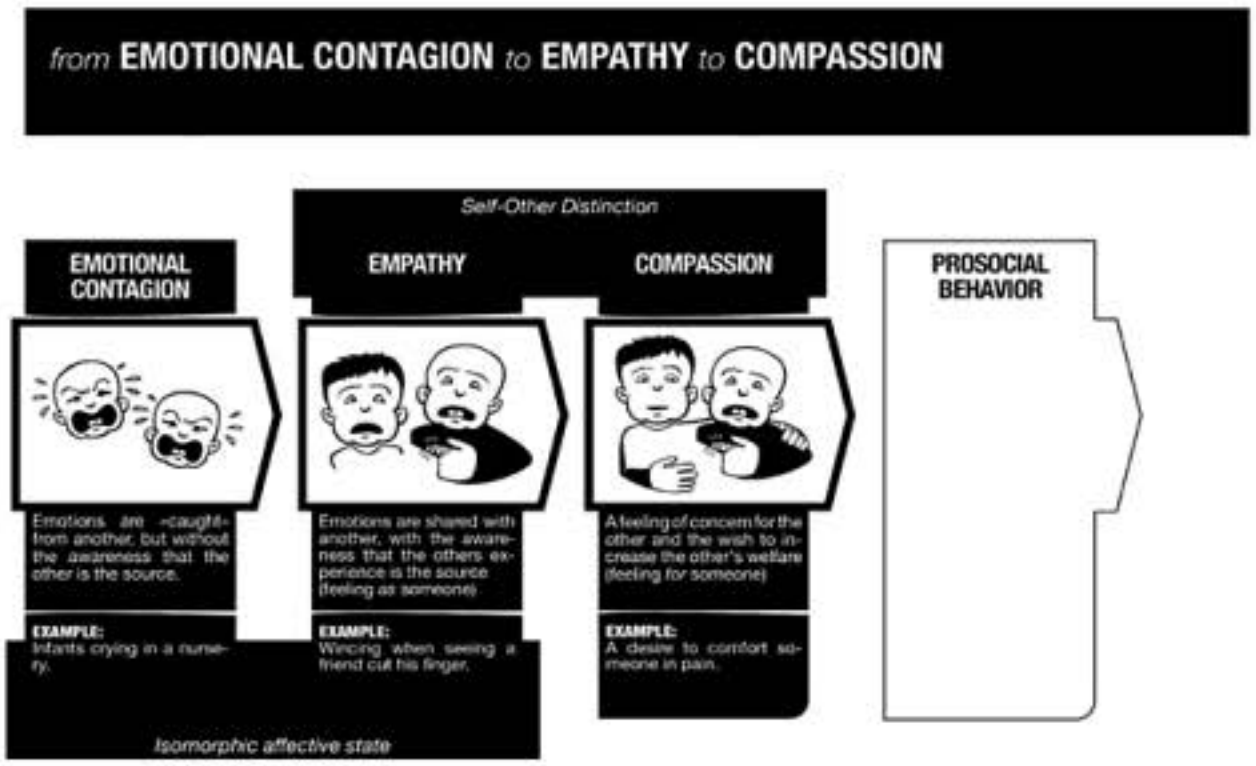

Figure 12.1 The conceptual relationship between emotional contagion, empathy, and compassion.

(de Vignemont & Singer, 2006; Hatfield, Cacioppo, & Rapson, 1994). The automatic spread of panic through a crowd or the collective crying among babies in a nursery are examples. The critical difference between empathy and emotional contagion is that empathy maintains a self-other distinction. In other words, it is clear to the empathizer that the target is the source of the affective state. Nevertheless, emotional contagion is likely a developmental precursor for empathy (Hatfield, Rapson, & Le, 2009; Singer & Lamm, 2009). Moreover, the mechanisms responsible for emotional contagion may also function in full-blown empathy, but with the fine-tuning of the self-other distinction (Decety & Jackson, 2004).

Sympathy and pity are also affective responses to another person's state, but without the isomorphic quality of empathy (Batson, 2009b). For example, "The fact that he's angry makes me sad," or "I'm happy that he's comfortable." Note that neither sympathy nor pity involves a clear element of emotional contagion. They are, however, likely to involve some degree of mentalizing as the empathizer uses various cues and beliefs about the other person's goals and experiences to infer an affective state. In terms of affect, sympathy is less direct than empathy (de Vignemont & Singer, 2006; Singer & Lamm, 2009); it involves feeling **for** someone, not feeling **as** someone (Batson, 2009b).

The term "compassion" is often used interchangeably with empathy. There is, however, an important distinction based on motivation and behavior. Compassion is characterized by the motivation to alleviate the distress of another (Baumeister & Vohs, 2007). Although empathy may allow a compassionate individual to know when and how to act (Batson, 2009a; Eisenberg, 2000), empathy does not always result in compassion. In fact, empathic distress may lead the empathizer to avoid the target individual (Batson, Fultz, & Schoenrade, 1987; Eisenberg & Fabes, 1990). Conversely, compassion is an approach-oriented response to the affective state of the other. It represents the prosocial consequences of empathic experience. One can also imagine antisocial responses that rely upon empathy (Singer & Lamm, 2009). Take, for example, a torturer who is uniquely skilled at knowing what will cause pain to his victims. The ability to experience vicarious pain would help him know how to hurt others.

The four parts of this definition of empathy provide a relatively well circumscribed territory for the study of this phenomenon. When it comes to neural underpinnings, we expect empathic

neural responses to represent affect (i) and to do so in a way that reflects the specific affective state (iii) of the empathized other (ii). At the same time, we expect distinctions between self and other representations indicative of the fact that the empathizer's feeling is vicarious and not direct (iv).

Neural representations of empathic states

In recent years, researchers have created a novel set of experimental paradigms to study the neuroscience of empathy. Initially this work was built on the premise that if empathy represents shared affect, then the neural representations of those vicarious states should show at least some overlap with self-generated representations of that same affective state (Avenanti, Bueti, Galati, & Aglioti, 2005; Botvinick, Jha, Bylsma, Fabian, Solomon, & Prkachin, 2005; Jackson, Meltzoff, & Decety, 2005; Keysers, Wicker, Gazzola, Anton, Fogassi, & Gallese, 2004; Morrison, Lloyd, Di Pellegrino, & Roberts, 2004;Singer, Seymour, O'Doherty, Kaube, Dolan, & Frith, 2004; Wicker, Keysers, Plailly, Royet, Gallese, & Rizzolatti, 2003). This "shared network hypothesis" emerged, in part, out of evidence for shared cognitive and neural representations of action and perception (Preston & de Waal, 2002; Gallese & Goldman, 1998; Prinz, 1997; Prinz, 2005). Research on the cognition mechanisms underlying action, for example, has consistently demonstrated that watching another person executing an action can interfere with the planning and execution of an incongruent action, and can facilitate the planning and execution of a congruent action. These types of findings suggest that a common coding exists for one's own actions and the perceived actions of others (Prinz, 2005; Prinz, 1997). This evidence was further bolstered by the discovery of mirror neurons in Macaque monkeys, neurons that respond to both action and the perception of action (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti & Craighero, 2004). Cognitive neuroscience, in turn, has demonstrated overlaps between regions representing one's own and others actions (Buccino, Binkofski, Fink, Fadiga, Fogassi, Gallese, et al., 2001; Jeannerod, 2001). Together these data led researchers to suggest that social cognition is built on the automatic simulation of others' behaviors (Gallese & Goldman, 1998; Keysers & Gazzola, 2007; Rizzolatti, Fogasi, & Gallese, 2001). The brain perceives others' actions and through their simulation infers the meanings of those actions.

Following this line of reasoning, the shared network hypothesis of empathy suggests that we understand others' affective states by recruiting the same networks that represent our own affective states. Direct and vicarious feeling rely on similar mechanisms.

Empathy for pain

While the shared network hypothesis has been tested in several affective domains, empathy for pain has been a particularly fruitful target. Pain lends itself to this line of research because it is easily manipulated and depicted within the laboratory. Both pain and empathizing for another person's pain are common and salient experiences. Perhaps most importantly, the "pain matrix", or networks responsible for representing pain (Apkarian, Bushnell, Treede, & Zubieta, 2005; Derbyshire, 2000; Peyron, Laurent, & Garcia-Larrea, 2000), is relatively well understood. As a consequence, researchers can make clear predictions about locations of overlapping representation and can theorize about the specific features of pain that are vicariously represented.

Social neuroscientists have used two distinct methods to manipulate and measure empathy for pain in the laboratory: picture-based and cue-based paradigms. In picture-based paradigms (e.g. Jackson et al., 2005), participants view pictures or videos depicting painful situations. For example, they might see a q-tip stroking a hand during a non-painful trial or a needle puncturing a hand during a painful trial (Lamm, Meltzoff, & Decety, 2010). Alternatively, the images can depict the

face of an individual as he or she experiences pain (Saarela, Hlushchuk, Williams, Schurmann, Kalso, & Hari, 2007). These studies allow researchers to measure neural responses while manipulating the nature of the vicarious stimulus, the location of that stimulus on the target's body, and the affective response of the pained individual.

Cue-based paradigms, on the other hand, use actual people instead of images as stimuli (Singer et al., 2004). During these experiments, multiple participants both receive and witness the delivery of painful stimuli (i.e. electric shocks to the hand). On each trial in such a study, a cue indicates (a) whether or not the stimulus for that trial will be painful, and (b) the recipient of that stimulus (i.e. self or other). Because this paradigm uses arbitrarily assigned cues to indicate trial type, any responses that emerge during other-recipient trials are entirely cue-triggered and cannot be caused by emotional contagion. They cannot be driven by simply seeing the recipient's body or by expressions of affect. In other words, empathic responses in these studies are the consequence of knowing the other person is in pain and imagining that state, not in perceiving the other person's actual response to that pain. The other important feature of this paradigm is that it includes both direct and vicarious pain trials; participants both experience and witness experience pain. As a consequence, researchers can perform a direct, within-subject comparison between a participant's own pain and his or her reaction to another's pain (see also Corradi-Dell'Acqua, Hofstetter, & Vuilleumier, 2011).

According to shared network hypotheses, empathy for another person's pain should activate components of the pain matrix. This activation should emerge when contrasting neural activity during trials depicting painful vs. non-painful trials. In experiments that include direct pain trials, one should also find overlap between self and other pain representations. Although cue-based and picture-based paradigms furnish distinct patterns of data, recent meta-analyses provide strong evidence for a core network for empathy for pain. One image-based meta-analysis representing 9 separate studies (Lamm, Decety, & Singer, 2011) and two coordinate-based meta-analyses representing 32 (Lamm et al., 2011) and 40 studies (Fan, Duncan, Greck, & Northoff, 2011) found significant bilateral anterior insula (AI), dorsal anterior cingulate cortex (ACC) and anterior midcingulate cortex (aMCC) activity during empathy for pain across a variety of experiments conducted by different research groups (Figure 12.2, Panel 1). Critically, these areas overlap with areas that emerged in a meta-analysis of activity during the direct experience of pain (Figure 12.2, Panel 4; Lamm et al., 2011).

Participant self-reports of empathic states and traits corroborate the role of these areas during the representation of vicarious pain. Activity in the ACC and left AI during other pain trials correlate (Singer et al., 2004; Singer, Seymour, O'Doherty, Stephan, Dolan, & Frith, 2006; Jackson et al., 2005; Lamm, Batson, & Decety, 2007a) with dispositional measures of empathy such as the Balanced Emotional Empathy Scale (Mehrabian & Epstein, 1972) and the Empathic Concern subscale of the Interpersonal Reactivity Index (Davis, 1983). Similar findings have been reported with the Empathy Quotient (Baron-Cohen & Wheelwright, 2004) and measures of emotional contagion (Lamm et al., 2007a; Doherty, 1997); see also (Jabbi, Swart, & Keysers, 2007). Reports of perceived target pain intensity or unpleasantness on a trial-by-trial basis also correlate with ACC and AI activity during those trials (Jackson et al., 2005; Saarela et al., 2007; Singer, Snozzi, Bird, Petrovic, Silani, Heinrichs, et al., 2008; Lamm, Nusbaum, Meltzoff, & Decety, 2007b; Cheng, Lin, Liu, Hsu, Lim, Hung, et al., 2007).

The core regions found across studies on empathy for pain map onto some, but not all, of the pain matrix. Here, qualitative distinctions between features of painful experience are critical. Specifically, the pain matrix can be divided into regions that represent sensory discriminative (the location of the pain, the quality of the nociceptive input, etc.) vs. affective and motivational components of pain (the experience of unpleasantness, avoidance motivation, etc.; Apkarian

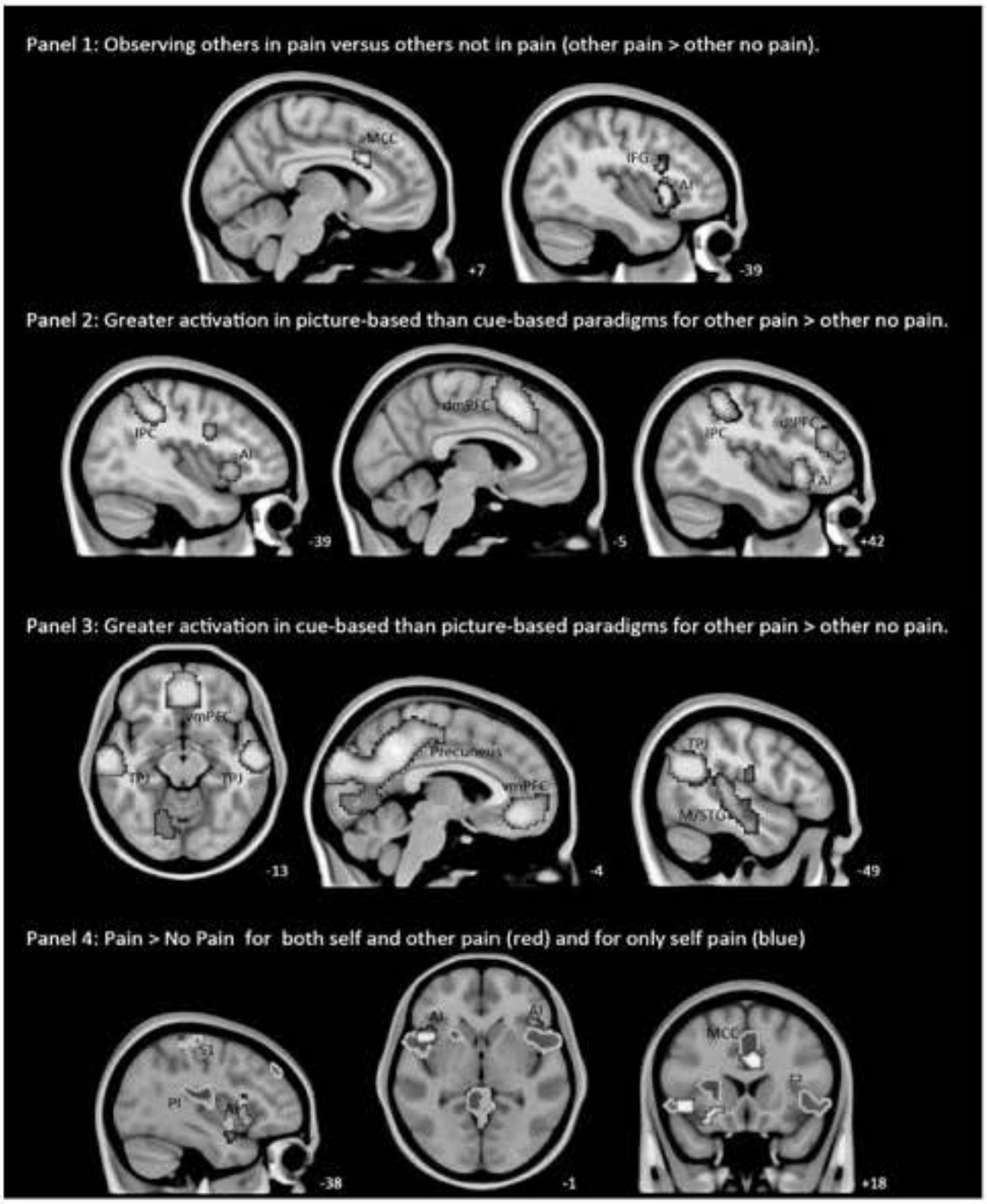

Figure 12.2 Results of a meta-analysis of empathy for pain studies (Lamm et al., 2011). The areas highlighted in Panel 1 showed significantly more activity when participants observed others in pain (as compared with trials in which they observed others not in pain). Panel 1 includes data from both cue-based and picture-based paradigms. Panel 2 depicts regions that showed higher activations for this contrast in picture-based paradigms. Conversely, Panel 3 depicts regions that showed higher activation for this contrast in cue-based paradigms. Panel 4 depicts areas that were common to both the experience of pain and the observation of others in pain (the bright spectrum) as well as areas that were unique to the direct experience of pain (the dark spectrum). See also Plate 5.

Analyses from 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): 2492–502. © 2011, Elsevier.

et al., 2005; Peyron et al., 2000). Somatosensory cortices (S1 and S2) and the posterior insula are implicated in the sensory discriminative components of pain (Apkarian et al., 2005; Maihöfner, Herzner, & Handwerker, 2006). In line with their role in pain localization, both the posterior insula and somatosensory cortices show activity contralateral to the location of the painful stimulus on

the body (Bingel, Quante, Knab, Bromm, Weiller, & Buchel, 2003; Brooks, Nurmikko, Bimson, Singh, & Roberts, 2002). Moreover, damage to the SI and SII (but not the AI or ACC) selectively impairs the ability to discern the quality and localization of pain without impairing the ability to experience an unpleasant feeling that varies with stimulus intensity (Ploner, Freund, & Schnitzler, 1999). Activation in S2, furthermore, varies with self reports of the sensory-discrimination, but not the other components of the pain (Maihöfner et al., 2006; Melzack, 1975).

On the other hand, the areas active across empathy for pain paradigms (the AI, dACC, and aMCC) have been implicated in pain's abstract and affective features (Price, 2000). AI activity, for example, varies not only with the level of noxious input, but additionally as a function of self-reported intensity (Kong, Gollub, Polich, Kirsch, Laviolette, Vangel, et al., 2008). Similarly, activity in the ACC correlates with self-reported unpleasantness of pain (Rainville, Duncan, Price, Carrier, & Bushnell, 1997). Not surprisingly then, vicarious pain involves feelings of discomfort and aversion without necessarily involving the more sensory-specific qualities of the stimulus. This is not to say, however, that areas associated with the sensory components of pain are never involved in empathic experience. Instead, it appears that different networks lead to the elicitation of the empathy and that the empathic experience of pain converges on this core network (Lamm et al., 2011).

Different routes to empathic experience

Comparisons between the results of cue-based and picture-based empathy for pain paradigms reveal important differences in patterns of activation. These differences suggest that the brain elicits empathic states through different computational routes (Decety & Hodges, 2006; Decety and Jackson, 2004; Singer, 2006). Along these lines, meta-analyses reveal that picture-based paradigms elicit activation in the anterior inferior parietal cortex (supramarginal gyrus and intraparietal sulcus) and ventral premotor areas (inferior frontal gyrus, pars opercularis; Figure 12.2, Panel 2; Lamm et al., 2011). Importantly, the joint activation of these two areas is also common to research on action observation (Van Overwalle & Baetens, 2009) and in the hearing or reading of sentences describing action (Aglioti & Pazzaglia, 2010). In fact, sequences of abstract non-biological stimuli (Schubotz & von Cramon, 2004) also activate this network, leading some to propose that it is involved in the prediction of external events (Schubotz, 2007). Given this account, the images in picture-based paradigms may set into motion a cascade of computations that ultimately provide predictive models of affective experience (Lamm et al., 2011). Participants attend to the picture of a knife pressing against a finger, these networks model the knife slicing through the skin that, in turn, elicits affective representations of the consequent laceration. Importantly, this cascade begins with the image of a body part and the implication of an event.

In the absence of images of the body, such a cascade of events is impossible. Instead, vicarious pain necessarily relies upon imagining the state of the other. Accordingly, cue-based vs. image-based events recruit more areas such as the precuneus, ventral parts of the medial prefrontal cortex, the posterior superior temporal cortex, the temporo-parietal junction, and the temporal poles (Figure 12.2, Panel 3;Lamm et al., 2011). These areas are traditionally implicated in Theory of Mind or mentalizing (Gallagher & Frith, 2003; Van Overwalle & Baetens, 2009). It is likely, then, that empathy in cue-based paradigms depends upon mentalizing processes. Participants imagine the condition of the other person and those processes, in turn, elicit empathic states.

The evidence that cue-based and picture-based paradigms elicit empathy via two different routes underscores the fact that understanding others relies on the activation of multiple different networks subserving social cognition (Singer, 2006). In everyday life, the brain uses bodily cues, symbols and pure imaginative inference to elicit empathic experience. While we consider empathic

states as distinctly affective in nature, the mechanisms that lead to empathic responses can rely on computations that would traditionally be labeled as cognitive. This amalgamation of processes allows us to feel empathy in both strictly symbolic circumstances (i.e. while reading a book) and more obviously visceral ones (i.e. while watching a boxing match). Moreover, the fact that various processes elicit empathy suggests that modulating empathic responses, learning empathic skills, and transforming empathy into prosocial behavior may involve a variety of different strategies that tap specific mechanisms.

Distinctions between direct and vicarious pain

One key component in our definition of empathy is that empathizing individuals share the affective state of a target, but preserve the distinction between self and other. The empathizer still identifies the target as the source of the experience. As such, one would expect differences between the neural representations of direct and vicarious pain. At the experiential level this point is obvious; watching a needle puncture someone else's skin can be distressing, but it's not the same feeling as getting pricked yourself. It's not surprising, then, that paradigms designed to directly compare self and vicarious pain find a host of activity that is unique to the direct experience of pain (Lamm et al., 2011).

In particular, cue-based paradigms (e.g. Singer et al., 2004) elicit strong activations in contralateral S1, posterior insula, and contralateral S2 during self pain, but no significant activity in these areas during the vicarious experience of pain. These findings again suggest that the sensory discriminative components of the pain matrix are not necessary for empathic experience. Although picture-based paradigms find higher activity in S1 and S2, those patterns often emerge ipsilaterally as well and on trials in which participants are exposed to both painful and non-painful stimuli (Lamm et al., 2011). Moreover, even patients with a congenital insensitivity to pain display significant activity in bilateral S1 when seeing pictures of others in pain (Danziger, Faillenot, & Peyron, 2009). Together these data argue against a specific mapping of the somatosensory features during vicarious pain. It is more likely that the activation in somatosensory cortices found in picture-based studies is a consequence of a more general activation elicited by the observation of touch on body parts (Keysers, Kaas, & Gazzola, 2010; Lamm et al., 2011).

Representations of direct and vicarious pain also appear to differ within the insula and cingulate cortex. While both activate anterior portions of the insula, direct pain uniquely activates its posterior subdivisions, which are associated more with sensory features of pain (Figure 12.2, Panel 4; Decety & Lamm, 2009; Lamm et al., 2011). Similarly, direct pain activates a larger portion of the cingulate cortex (Lamm et al., 2011) with distinct activation patterns (Decety & Lamm, 2009; Morrison & Downing, 2007). Connectivity analyses furthermore suggest that overlapping regions responsible for both self and vicarious affect are embedded in larger and divergent networks (Jabbi, Bastiaansen, & Keysers, 2008; Zaki, Ochsner, Hanelin, Wager, & Mackey, 2007).

While it is clear that activation patterns are distinct for direct and vicarious pain, their significant overlap in areas critical for affective experience supports the claim that they rely upon some of the same computations. The spatial resolution of fMRI, however, prevents us from determining whether or not the two states activate the same subpopulations of neurons within the overlapping voxels. Except for one subject in one single cell recording study (Hutchison, Davis, Lozano, Tasker, & Dostrovsky, 1999), there is no direct evidence for precise neuronal overlap between direct and vicarious pain. Nevertheless, recent work using multivariate pattern analysis of fMRI data provides the strongest evidence yet (Corradi-Dell'Acqua et al., 2011). This study looked at multivoxel patterns of activity during direct thermal pain to the hand and the observation of hands in painful situations. Whole brain analyses revealed similar patterns in the AI (bilaterally) in the

two conditions. Region of interest analyses, furthermore, found overlap in the middle insula and middle cingulate cortex. The fact that distributed ensembles of voxels and not simply isolated voxels showed common patterns provides powerful evidence for the shared network hypothesis of empathy for pain.

The role of the anterior insula and cingulate cortex in empathy

The core network found in empathy for pain (the AI and dACC/aMCC) also emerges in research on empathy for other forms of affect. For example, participants in a study on disgust (Jabbi et al., 2008) either tasted a bitter liquid, watched videos of actors tasting bitter liquids, or imagined doing so. All three scenarios elicited activity in the AI and adjacent frontal operculum. Similarly, both inhaling disgusting odorants and seeing faces expressing disgust activated portions of the ACC and AI (Wicker et al., 2003). These regions further emerged in studies on empathic responses to bodies expressing fear (Gelder, Snyder, Greve, Gerard, & Hadjikhani, 2004) and anger (Grosbras & Paus, 2005). Even the sweat of anxious individuals triggered activity in these areas (Prehn-Kristensen, Wiesner, Bergmann, Wolff, & Jansen, 2009). Evidence also suggests that these areas are involved in representing vicarious responses to more obviously social experiences. Specifically, they emerged when participants were exposed to scenes in which targets were embarrassed (Krach, Cohrs, Loebell, Kircher, Sommer, Jansen, et al., 2011) or socially excluded (Masten, Eisenberger, Borofsky, Pfeifer, McNealy, Mazziotta, et al., 2009). Together these data suggest that the AI and dACC/aMCC comprise a network for a multitude of empathic experiences (Bernhardt & Singer, 2012). Given that, what computations occur in these areas and how do they work together to produce empathic experience?

The anterior insula

The insula has long been associated with interoception (Craig, 2002). Functional neuroimaging studies demonstrate its involvement in a wide variety of visceral representations including thirst, bladder distension, sexual arousal, temperature perception, disgust, autonomic arousal, and (of course) pain (Craig, 2009). The AI, specifically, is implicated in the conscious perception of internal states (Craig, 2009), its engagement correlating with interoceptive abilities (Critchley, 2005). For example, AI activity predicts accuracy on a heartbeat detection task in which one compares one's own heartbeat to external feedback. Both performance on this task and self-reports of visceral awareness also correlate with the cortical thickness of the AI (Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004).

Damasio famously linked the insula's bodily associations with emotional experience. According to his influential model (Damasio, 1994), this region integrates visceral and sensory signals and, in doing so, produces emotional experience. The link between interoceptive awareness and emotional experience has, indeed, been supported by empirical data (Barrett, Quigley, Bliss-Moreau, & Aronson, 2004; Pollatos, Gramann, & Schandry, 2007). Craig (2002, 2009) has further proposed that bodily states are initially represented in the posterior or mid insula, and are then remapped in the AI where they contribute to consciously accessible feeling states. In the domain of pain, these claims are further supported by recent work using direct electrical stimulation of the cortical surface during presurgery evaluations of patients with epilepsy (Mazzola, Isnard, Peyron, & Mauguiere, 2011). In over 4000 cortical stimulations of 164 patients, only stimulation of the posterior insula and medial parietal operculum elicited pain responses. Connectivity patterns within the insula and between the insula and other structures further support a posterior-to-anterior mapping of visceral input to conscious and affective remapping (Craig, 2002, 2009).

With its dense connections to both limbic and forebrain regions (Craig, 2009; Critchley et al., 2004; Kurth, Zilles, Fox, Laird, & Eickhoff, 2010), the AI is ideally situated to be a conduit between bodily states and more conscious emotional experience. Accordingly, the AI emerges in multiple studies in which participants focus on their feelings. For example, AI emerged when participants attended to joyful voices (Johnstone, Reekum, Oakes, & Davidson, 2006), or read a sentence expressing a joyful feeling and imagined themselves feeling that joy (Takahashi, Matsuura, Koeda, Yahata, Suhara, Kato, et al., 2008). The role of the AI in affective experience is nicely illustrated by research on alexithymia, a subclinical phenomenon in which individuals have difficulty identifying and describing their emotions. In one study, participants completed a task in which they were exposed to a series of images. Their task on each trial was to either rate their emotional reaction to the image (to introspect) or to simply judge the color balance of the image. Alexithymics showed relatively reduced AI activity when introspecting about their emotional responses to unpleasant images (Silani, Bird, Brindley, Singer, Frith, & Frith, 2008). Similarly, alexithymics showed reduced empathic responses in anterior insula when perceiving close others in pain (Bird, Silani, Brindley, White, Frith, & Singer, 2010).

The AI is probably involved not only in the conscious representation of affective states, but also in computations of prediction and prediction error (Paulus & Stein, 2006; Singer, Critchley, & Preuschoff, 2009). In one study on the anticipation of pain participants completed a series of trials in which they either received painful or non-painful stimulation (Ploghaus, Tracey, Gati, Clare, Menon, Matthews, et al., 1999). Before each trial, the type of stimulation was revealed via a colored light. Participants showed significant AI activity when they saw the pain cue, but before the delivery of the pain, indicating a representation of anticipation of the painful shock. Based on these and other data, researchers have proposed that AI computes predication error between anticipated states and actual visceral input (Paulus & Stein, 2006; Singer et al., 2009). These affective predictions have two critical consequences (Singer et al., 2004; Singer et al., 2009). First, they allow us to anticipate our physiological reactions to emotional stimuli. Secondly, they simulate the affective reactions of other people (i.e. vicarious pain).

Neuroeconomics research further implicates the AI in processing and prediction of risk and uncertainty (Critchley, Mathias, & Dolan, 2001; Grinband, Hirsch, & Ferrera, 2006; Paulus, RogalskY, Simmons, Feinstein, & Stein, 2003; Preuschoff, Quartz, & Bossaerts, 2008). The AI is active during tasks which are risky, ambiguous, or complex (Elliott, Friston, & Dolan, 2000; Grinband et al., 2006; Huettel, Stowe, Gordon, Warner, & Platt, 2006). These data suggest that the AI predicts risk and uncertainty and computes errors between those predictions and actual outcomes (Singer et al., 2009). For example, bilateral AI activity emerged when participants waited for the outcome of a risky decision and the level of activity reflected the risk prediction error once the outcome was known (Preuschoff et al., 2008).

Based on these findings and on the involvement of the AI in representing direct and vicarious feeling states, Singer and colleagues have suggested a broader model of AI functioning (Singer et al., 2009). Within this model the AI integrates information about online and projected feeling states. It processes incoming sensory, bodily, and contextual information, while generating predictions for the affective consequences of anticipated events. By comparing these two channels of data, it calculates and refines estimates of outcomes, uncertainty, and their prediction errors. Together these functions produce a global feeling state, which reflects the integration of interoception, prediction, and risk. Critically, this integration would allow the AI to drive emotional learning and decision making. When considered in terms of empathy, the AI may compute the projected feelings states of another person and may, furthermore, compare those states with online information (e.g. facial or vocal expressions, bodily state, and etc.). Such computations would allow us to learn from others'

positive or negative experiences and to, perhaps, provide help or support when they are needed. In other words, we can learn and make decisions, not only from our own emotional states, but from the observed or imagined states of others.

The cingulate cortex

The cingulate cortex also emerges across studies on empathy, specifically the dACC/aMCC. Functionally, this region has been implicated in a wide variety of phenomena. A recent meta-analysis of 939 studies found that overlapping portions of the dACC/aMCC are involved in representing negative affect, pain, and cognitive control (Shackman, Salomons, Slagter, Fox, Winter, & Davidson, 2011). Other data also implicate the regions in response selection (Medford & Critchley, 2010). Researchers consistently find concurrent activation in the AI and these regions of the cingulate, particularly in emotion-related paradigms (Craig, 2009; Medford & Critchley, 2010). In line with those findings, resting state fMRI connectivity analyses show a close functional relationship between the AI and these areas (Taylor, Seminowicz, & Davis, 2009; Harrison, Pujol, Ortiz, Fornito, Pantelis, & Yucel, 2008), a relationship that is supported by dense anatomical interconnections (Bernhardt & Singer, 2012; Medford & Critchley, 2010).

As mentioned above, these portions of the ACC contribute to the affective component of the pain matrix (Apkarian et al., 2005; Rainville et al., 1997). Specifically, they're associated with the motivational component of the response and likely play a critical role in preparing responsive action (Morrison & Downing, 2007; Vogt & Sikes, 2009). In line with this account, the dACC/aMCC receives direct projections from pain pathways; caudal divisions of the cingulate near the dACC/aMCC also have strong functional connections to sensorimotor regions (Margulies, Kelly, Uddin, Biswal, Castellanos, & Milham, 2007). Research on animal models provides evidence that these regions are involved in motivated action. Ablation of the ACC in rats selectively reduces avoidant behavior without reducing sensitivity to noxious stimuli (LaGraize Labuda, Rutledge, Jackson, & Fuchs, 2004). Furthermore, single neuron recordings in monkeys have identified neurons in the ACC that selectively fire in response to cues for forthcoming pain or reward stimuli that they can either approach or avoid (Koyama, Keichiro, Tanaka, & Mikami, 2001).

The joint activation of AI and regions of the ACC also emerges frequently in studies on emotion (Craig, 2009; Medford & Critchley, 2010). As with pain, dACC/aMCC activity here may represent the mapping of affective responses into motivational and somatic domains (Craig, 2009; Pollatos et al., 2007; Medford & Critchley, 2010). Accordingly, joint ACC and AI activity has been shown in response to emotional facial expressions and, more to the point, is associated with heart rate changes (Critchley, Rotshtein, Nagai, O'Doherty, Mathias, & Dolan, 2005a). Further evidence links aMCC activity more generally with autonomic arousal (Critchley, Tang, Glaser, Butterworth, & Dolan, 2005b). Within the domain of empathy, the link between the AI and the dACC/aMCC likely represents the causal chain from projections of the target's feeling state to motivational and behavioral responses. In other words, feeling that others are in distress can drive us to flinch, cringe, or act. Of course the nature of the behavioral response varies. We will return to the behavioral consequences of empathy at the end of this chapter.

Empathy in somatosensation

Although we have focused on the role of the AI and ACC in empathy, research also demonstrates empathy for touch in regions more directly associated with somatosensation (Blakemore, Bristow, Bird, Frith, & Ward, 2005; Ebisch, Perrucci, Ferretti, Gratta, Romani, & Gallese, 2008; Keysers et al., 2004, 2010; see also Keysers Thioux, and Gazzola, this volume). Participants have consistently shown common activation in the secondary somatosensory cortex when they both experience

and observe touch. In one fMRI study, participants watched videos of legs being stroked, or had their own legs stroked in a similar fashion. This overlap in activation within in the secondary somatosensory cortex in the two conditions (Keysers et al., 2004) has been replicated in replicated in subsequent research (Ebisch et al., 2008; Schaefer, Xu, Flor, & Cohen, 2009). One study, furthermore, identified an individual who experiences conscious tactile sensation when watching another person being touched (Ebisch et al., 2008). While both she and normal controls showed responses to observed touch in the somatosensory cortices, the activity was significantly greater for this individual. Together these data demonstrate that the observation and experience of somatosensation recruits common networks.

The modulation of empathic responses

It is doubtful that anyone feels empathy for all people at all times. In fact, it is quite easy to come up with situations in which we feel more or less empathy toward an individual based on who that person is or how they have behaved. There are also clear differences between individuals in the ability or motivation to empathize with others. Given this wide variability, understanding the modulation of empathy is critical to understanding the phenomenon itself. Along these lines, researchers have explored the effects of context, interpersonal factors, and individual empathic capacities to better understand how and why empathic responses vary (see also de Vignemont and Singer, 2006; Hein & Singer, 2008).

Our ability and willingness to empathize with others is strongly affected by their identity and behavior. While the original cue-based, empathy-for-pain paradigm described above looked at empathy for loved ones (Singer et al., 2004), follow-up research has used the same paradigm to examine empathy for strangers. In one such study (Singer et al., 2006), participants came into the lab and interacted with confederates whom they believed to be other participants and who differed in their fair or unfair behavior toward the participant. They first completed an economic game with the confederates. During the game, one confederate played fairly and the other participant played unfairly, "defecting" in economic exchanges by failing to reciprocate the participant's offers. Later, the participant and confederates completed the empathy-for-pain task. When fair players received painful shocks, participants exhibited the same empathic response exhibited in the prior study (i.e. activity in the AI and ACC). When unfair players received shocks, however, male participants showed relatively reduced responses in these regions. Moreover, they exhibited increased activity in areas that have typically been associated with reward processing (i.e. the nucleus accumbens). These increases also correlated with the self-reported desire for revenge.

This effect was replicated in a subsequent study in which participants witnessed both ingroup and outgroup members receiving a painful stimulus (Hein, Silani, Preuschoff, Batson, & Singer, 2010). Participants were soccer fans and they interacted with fellow fans of their favorite team (in-group members) and fans of their favorite team's rival (out-group members). The results showed stronger responses in the left AI when participants witnessed an in-group member vs. an out-group member suffer. As with unfair players in the earlier study, witnessing out-group members elicited activation in the nucleus accumbens that was modulated by group perception.

Characteristics of a target person's perceived affective state can also moderate the empathic response. In one study (Saarela et al., 2007), participants were shown photos of faces of chronic pain patients who were experiencing varying levels of acute pain. Participants showed more activity in several areas including the AI and ACC when exposed to the acute pain photos. Moreover, their estimates of targets' pain intensity correlated with the strength of activation in these areas (left ACC, left inferior parietal lobule (IPL), and bilateral AI). Here, the affective facial cues modulated the empathic responses, even in the absence of bodily cues about the painful stimulus itself.

Conversely, contextual information can alter empathic responses to identical images of the body. In one study, participants were shown similar images of hands undergoing medical procedures (Lamm et al., 2007b). One image type depicted a painless biopsy performed on an anesthetized hand while the other depicted a painful injection into a hand. Despite the relatively abstract information about the nature of the photos, participants showed reduced empathic responses in the AI and aMCC when exposed to the anesthetized vs. non-anesthetized hand. Knowing the consequences of a painful event can also affect the empathic neural response. For example, when participants watched videos of a painful procedure, they showed weaker activity in the aMCC and AI when they believed that procedure to be therapeutically effective than when they believed it to be ineffective.

Attention and imagination also play critical parts in the modulation of empathy. Participants who observed images of hands in painful situations showed stronger activation in the AI and ACC when they focused on the intensity of the person's pain as opposed to physical features of the image (Gu & Han, 2007). Similarly, perspective-taking can alter the neural response. Participants imagining themselves in a painful situation vs. imagining another person in that situation show enhanced responses to the images, notably in the insula and aMCC (Jackson et al., 2006; Lamm et al., 2007a).

As mentioned in the previous section, individual differences in self-report measures of empathy (Baron-Cohen & Wheelwright, 2004; Davis, 1983; Mehrabian & Epstein, 1972) correlate with empathic responses to equivalent stimuli. Conversely, empathic deficits, on the other hand, emerge in various clinical phenomena. Of course the most obvious disorder for which we would expect deficits is psychopathy. While there is not yet direct evidence for a reduced empathic neural response in psychopaths, less AI grey matter volume has been associated with weaker empathy scores in adolescents with conduct disorder (Sterzer, Stadler, Poustka, & Kleinschmidt, 2007). Research on adult psychopaths, furthermore, has shown reduced activity of the amygdala and AI during the anticipation of pain (Birbaumer, Veit, Lotze, Erb, & Hermann, 2005). More data are necessary to make a claim about the neural nature of empathic deficits in psychopathy.

The data on empathic deficits in alexithymics are clearer. As mentioned above, alexithymics have reduced introspective abilities which appear to translate into reduced empathic responses (Silani et al., 2008). Silani and colleagues' findings have been replicated in subsequent research (Bird et al., 2010). Again, empathic neural activation elicited by the pain of a close other was modulated by individual levels of alexithymia. Importantly, this sample included individuals with autism spectrum conditions. When analyses accounted for levels of alexithymia, empathic responses were comparable between autistic and control groups. As such, although alexithymia and autism spectrum disorders show high comorbidity, there is no necessary deficit in empathy in autism. This double dissociation further underscores the distinction between empathic and mentalizing abilities as autism spectrum disorders are known to be associated with severe theory of mind deficits (Baron-Cohen, 1995).

While we commonly consider empathy to be a positive trait, there are some domains in which a controlled empathic response is clearly beneficial. Health practitioners, for example, would have a terrible time if they winced or cringed every time they had to perform a painful procedure. One study by Cheng and colleagues (2007) addressed this point, exposing both acupuncturists and laymen to images of needles being inserted into different parts of the body. As predicted, only the laymen showed neural activation characteristic of empathic responses. Of course, is seems likely that while acupuncturists may control their empathic responses to pain, they likely preserve empathic responses in other domains. After all, different circumstances require different responses. Along those lines, the complex relationship between empathy and behavior is the topic of our final section.

From empathy to prosocial behavior

Although our definition of empathy does not refer directly to behavior, one would expect such a basic component of social life to influence it. Indeed, the network of regions we've focused on here suggest a causal pathway from other-oriented prediction to conscious feeling state to motivation (Bernhardt & Singer, 2012; Craig, 2009). As such, empathy likely prepares one to respond and, possibly, to act. At the individualistic level, sharing other peoples' feelings allows us react to their distress or joy so that we can avoid their mistakes or emulate their successes. In more prosocial terms, empathy allows us to respond to the needs of distressed others or to share in their joy.

These two putative functions of empathy can imply very different behavioral consequences. Accordingly, empathy researchers have long drawn a key distinction between two different empathic reactions: empathic concern and empathic distress (Batson, 2009a; Eisenberg, 2000; Klimecki & Singer, 2012). Empathic concern is akin to sympathy. The concerned individual responds to the distressed state of another, but with an approach motivation—they feel a desire to care for the target. Empathic distress, on the other hand, is an aversive state associated with avoidance motivation. The empathically distressed individual assumes the distressed feelings of the target to such an extent that they must physically or symbolically flee the situation. They are incapable of helping.

Social neuroscience is only beginning to explore this difference and to better understand how empathy might lead to the kind of approach behaviors associated with helping. Lamm and colleagues (2007a), for example, point to the fact that when participants consider a painful scene using self (vs. other) perspective-taking, they show a stronger activation in components of the core empathy network (the insula and the aMCC) and in the amygdala (which among other things plays a critical role in fear-related behaviors; LeDoux, 2003). The assumption of the first person perspective here may push the experience into empathic distress such that the individual experiences the kinds of personal distress and avoidance motivation associated with direct pain.

On the other hand, Hein and colleagues (2010) have provided neural evidence that empathy can motivate costly helping. As part of the abovementioned study on empathy for in-group vs. out-group members, participants had the opportunity to receive a painful stimulus in order to reduce the painful stimulus delivered to another player. Participants who showed more AI activation while seeing an in-group member suffer were more likely to help that person. Conversely, participants who showed more nucleus accumbens activity (associated with reward), while seeing the out-group member suffer, were less likely to help.

Work on social exclusion extends these findings (Masten, Morelli, & Eisenberger, 2010). During this fMRI experiment, participants observed one person being excluded by two other people during a computerized ball-tossing game (Williams et al., 2000). After the scanning period, participants were asked to send emails to the players whom they had observed. Coders rated the degree to which the emails sent to the ostracized individual were comforting, supportive, and attempted to be helpful. Analyses revealed positive relationships between these prosocial communications and activity in the right AI and the medial prefrontal cortex during the exclusionary event (Masten et al., 2010; see also Mathur, Harada, Lipke, & Chiao, 2010). Given these data, it appears that empathic experience, and its neural components, can indeed promote helping behaviors.

Outlook

A crucial question for the future study of empathy is how and why vicarious feelings sometimes lead to empathic distress and avoidance, and other times lead to empathic concern and helping. Emotion regulation likely plays an important role in these processes, but which specific regulatory

strategies are effective, the nature of their neural representations, and how they interact with neural representations of empathy remain to be seen.

Important questions also remain regarding the relationship between empathy and different psychopathologies. For example, how does empathy manifest (or fail to manifest) in individuals with depression, borderline personality disorder, or narcissistic personality disorder (see Ritter, Dziobek, Preissler, Rütter, Vater, & Fydrich, 2011)? It will also be critical to disentangle ways in which mentalizing pathways and empathy pathways are differentially affected in these disorders. Besides the obvious application in clinical domains, this line of research will help distinguish between the various mechanisms that drive social cognition.

The plasticity of empathy is another key frontier. Can one be trained to be more empathic or to better transform the empathic response into prosocial action? If so, what are the components of effective empathy training? Along similar lines, neuroscience is only beginning to investigate the developmental trajectory of empathy (e.g. Decety, Michalska, & Akitsuki, 2008) and, more generally, social emotions. Future work in this domain will help identify the critical periods in which social emotions emerge, and the factors that facilitate their emergence.

Ideally, these new lines of inquiry will translate the basic findings from the neuroscience of empathy into everyday benefits. Empathy, after all, is one of human nature's more appealing traits.

Acknowledgements

We would like to thank Ralf Hartmann for his assistance in creating Figure 12.1. We also thank Claus Lamm and Boris Bernhardt for their help with Figure 12.2.

References

Aglioti, S.M., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale* **206**: 141–51.

Apkarian, A.V., Bushnell, M.C., Treede, R.D., & Zubieta, J.K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain* **9**: 463–84.

Avenanti, A., Bueti, D., Galati, G., & Aglioti, S.M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature neuroscience* **8**: 955–60.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press/ Bradford Books.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism Development Disorders* **34**: 163–75.

Barrett, L.F., Quigley, K.S., Bliss-Moreau, E., & Aronson, K. R. (2004). Interoceptive sensitivity and self-reports of emotional experience. *Journal of Personal Social Psychology* **87**: 684–97.

Batson, C. D. (2009a). Empathy-induced altruistic motivation. In: M. Mikulincer & P. R. Shaver (Eds), *Prosocial Motives, Emotions, and Behavior* (pp. 15–34). Washington, D.C.: American Psychological Association.

Batson, C. D. (2009b). These things called empathy: eight related, but distinct phenomena. In: J. Decety & W. Ickes (Eds), *The Social Neuroscience of Empathy* (pp. 3–15). Cambridge: MIT Press.

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

Baumeister, R.F. & Vohs, K.D. (2007). *Encyclopedia of Social Psychology*. Thousand Oaks: Sage Publications.

Bernhardt, B. & Singer, T. (2012). The neural basis of empathy. *Annual Reviews of Neuroscience* **35**: 1–23.

Bingel, U., Quante, M., Knab, R., Bromm, B., Weiller, C. & Buchel, C. (2003). Single trial fMRI reveals significant contralateral bias in responses to laser pain within thalamus and somatosensory cortices. *NeuroImage* **18**: 740–8.

Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., & Flor, H. (2005). Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. *Archives of General Psychiatry* **62**: 799–805.

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: Journal of Neurology* **133**: 1515–25.

Blakemore, S.J., Bristow, D., Bird, G., Frith, C., & Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision-touch synaesthesia. *Brain* **128**: 1571–83.

Botvinick, M., Jha, A.P., Bylsma, L.M., Fabian, S.A., Solomon, P.E., & Prkachin, K.M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage* **25**: 312–19.

Brooks, J.C., Nurmikko, T.J., Bimson, W.E., Singh, K.D., & Roberts, N. (2002). fMRI of thermal pain: effects of stimulus laterality and attention. *NeuroImage* **15**: 293–301.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G. & Freund, H.J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience* **13**: 400–4.

Cheng, Y., Lin, C.P., Liu, H.L., Hsu, Y.Y., Lim, K.E., Hung, D., & Decety, J. (2007). Expertise modulates the perception of pain in others. *Current Biology* **17**: 1708–13.

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. *Journal of Neuroscience* **31: 17966–8006.**

Craig, A.D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *National Review of Neuroscience* **3**: 655–66.

Craig, A.D. (2009). How do you feel—now? The anterior insula and human awareness. *National Review of Neuroscience* **10**: 59–70.

Critchley, H.D. (2005). Neural mechanisms of autonomic, affective, and cognitive integration. *Journal of Comprehensive Neurology* **493**: 154–66.

Critchley, H.D., Mathias, C.J., & Dolan, R.J. (2001). Neural activity in the human brain relating to uncertainty and arousal during anticipation. *Neuron* **29**: 537–45.

Critchley, H.D., Rotshtein, P., Nagai, Y., O'Doherty, J., Mathias, C.J., & Dolan, R.J. (2005a). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *NeuroImage* **24**: 751–62.

Critchley, H.D., Tang, J., Glaser, D., Butterworth, B., & Dolan, R.J. (2005b). Anterior cingulate activity during error and autonomic response. *NeuroImage* **27**: 885–95.

Critchley, H.D., Wiens, S., Rotshtein, P., Ohman, A. & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience* **7**: 189–95.

Damasio, A.R. (1994). Descartes' error and the future of human life. *Scientific American* **271:** 144.

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**, 203–12.

Davis, M.H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology* **44:** 113–26.

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences* **10**: 435–41.

Decety, J., & Hodges, S.D. (2006). The social neuroscience of empathy. In: P. A. M. Lange (Ed.), *Bridging Social Psychology Benefits of Transdisciplinary Approaches* (pp. 103–10). Mahwah: Erlbaum.

Decety, J. & Jackson, P.L. (2004). The functional architecture of human empathy. *Behavioral Cognitive Neuroscience Review* **3**: 71–100.