The Self to Other Model of Empathy:
Providing a New Framework for Understanding Empathy Impairments in Psychopathy, Autism, and Alexithymia

Geoffrey Bird\textsuperscript{1,2} and Essi Viding\textsuperscript{2,3}

\textsuperscript{1}MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, DeCrespigny Park, London, SE5 8AF, UK

\textsuperscript{2}Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N 3AR, UK

\textsuperscript{3}Department of Clinical, Educational and Health Psychology, Division of Psychology and Language Sciences, University College London, Gower Street, London, WC1E 6BT

Correspondence can be addressed to either author.
GB: Email: geoff.bird@kcl.ac.uk Fax: +44 (0) 20 7848 0866
EV: Email: e.viding@ucl.ac.uk Fax: +44 (0) 20 7679 2000
Abstract

Despite increasing empirical and theoretical work on empathy, particularly on the content of empathic representations, there is a relative lack of consensus regarding the information processing necessary for empathy to occur. Here we attempt to delineate a mechanistic cognitive model of empathy in order to provide a framework within which neuroimaging work on empathy can be located, and which may be used in order to understand various disorders characterised by atypical levels of empathy. To this end data from individuals with psychopathy, autism, and alexithymia inform the model, and the model is used to provide a unifying framework for any empathy impairments seen in these disorders. The model adopts a developmental framework and tries to address the four difficult questions of empathy: How do we know what another is feeling? What is the role of theory of mind in empathy? How does the state of another cause a corresponding state in the self? How do we represent another’s emotion once emotional contagion has taken place?

Keywords: Empathy, Psychopathy, Autism, Alexithymia, Theory of Mind, Emotional Contagion, Emotion Recognition.
Introduction

Empathy is a basic, highly-valued human characteristic. The presence of empathy in typical individuals, and its absence in a number of psychopathologies, has attracted a great deal of research interest (e.g. Blair, 2005; de Vignemont & Singer, 2006). Despite this interest there is a relative lack of consensus regarding the information processing steps necessary for the instantiation of empathy (Singer & Lamm, 2009).

Hoffman (2000) defines empathy as, “any process where the attended perception of the object’s state generates a state in the subject that is more applicable to the object’s state or situation than to the subject’s own prior state or situation.” This characterisation of a shared (i.e. isomorphic) feeling state with an individual who is the object of empathy is relatively uncontroversial and would likely be accepted as a broad descriptive definition of empathy by most researchers/theorists working in the field. Using this definition empathy research can be ‘unpacked’ in two different ways: 1) The study of outcomes, i.e. the degree of congruence of feeling states between the self and another (What is shared?); 2) The study of mechanisms, i.e. what features of the cognitive and affective systems enable empathy to be experienced (How is it shared?).

Recent empirical and theoretical contributions have particularly advanced the study of ‘What is shared?’ (e.g. Avenanti et al., 2006; Silani et al., 2013; Singer et al., 2004; Wicker et al., 2003; see Lamm et al., 2010 for a review). Preston and de Waal (2002) proposed that observing or imagining another person in a particular emotional state automatically activates a representation of that state (and accompanying autonomic and somatic responses) in oneself. Several studies support this position by demonstrating that observation of another’s emotional or sensory state activates a subset of those
neural areas that are active when the state is experienced by the self (Singer et al., 2004; Wicker et al., 2003; Keysers et al., 2004; Blakemore et al., 2005). For example, Singer et al. (2004) showed that affective, but not somatosensory, parts of the self-pain network were active in response to knowledge that a loved one was experiencing pain; while observing body parts in painful situations is associated with activity also in somatosensory areas of the self-pain network (Lamm et al., 2011).

It is important to distinguish the empathic process from a number of other socio-cognitive processes that also involve the representation of another’s state (Shamay-Tsoory, 2014). These processes can be differentiated on the basis of the information that is represented and on the neural networks upon which the processes rely. Empathy involves the representation and sharing of affective information and is thought to rely on the anterior insula and anterior cingulate cortex (Lamm et al., 2011). ‘Theory of Mind’ (or mentalising) entails representation of mental states and is thought to rely on a network of brain regions including the temporo-parietal junction (TPJ) and medial prefrontal cortex (MPFC) (Frith and Frith, 2003), while imitation (or mimicry) concerns representation and sharing of action plans and relies on activity in inferior frontal gyrus and inferior parietal cortex (Catmur et al., 2009; Heiser et al., 2003).

de Vignemont and Singer (2006) have argued that knowing ‘What is shared’ is also necessary in order to distinguish empathy from sympathy. They specify that empathy and sympathy can be differentiated on the basis of the congruence between the affective states of the other and of the self and propose that empathy is demonstrated only when the affective states of the empathiser and the object of empathy are isomorphic. In addition, they argue that this initial ‘emotional contagion’ needs to be accompanied by the knowledge that the source of one’s own emotional state is the
other. Under this account, ‘What is shared’ are specific affective states and the empathiser recognises that sharing has taken place.

Despite the recent progress in characterising what is shared there has been less progress made in constructing an information processing model of empathy (with some exceptions discussed below), i.e. in establishing ‘How is it shared?’ The purpose of this article is to provide, based in part upon previous work in this area (Blair, 2005; Lamm et al., 2011), a model of the mechanisms by which the affective state of another may cause an empathic state in the self. Setting out this model will enable us to locate more precisely the source of empathy deficits in different conditions that appear to have at least partially overlapping behavioural manifestations; psychopathy, Autism Spectrum Disorders (henceforth ‘autism’), and alexithymia. To this end, existing data from individuals with these conditions inform the model and we also suggest ways in which the model can elucidate why empathy deficits may be seen in these conditions. The model may also provide an additional framework for interpreting the burgeoning neuroimaging literature on empathy.

There are a number of difficult mechanistic questions that must be addressed by any cognitive model of empathy:

1) How do we know what another is feeling? For empathy to have occurred, the empathiser must be in the same affective state as the object of empathy. It therefore follows that a necessary first step in the generation of that matching state should be the accurate classification of another’s affective state.
2) What is the role of Theory of Mind (ToM) in empathy? Is the representation of another’s affective state (required for empathy to occur) reliant upon a system to represent their mental state?

3) How does emotional contagion occur? In other words, how does another person’s emotion elicit a matching emotion in the self? We suggest that this is perhaps the most crucial question with respect to empathy, but has received the least attention.

4) How do we represent another’s emotion once emotional contagion has taken place? In order for empathy to have occurred, the empathiser has to be in the same affective state as the object of empathy. Thus, the empathiser will receive a number of veridical psychophysiological cues about their own emotional state. Despite this reliable indicator that the emotional state ‘belongs to the self’, i.e. that it is the empathizer’s own state perceived in the first-person, in order for emotional contagion to develop into empathy the affective state of the self must also be assigned to the other. Most theories of empathy seek to explain how the emotional state of the other is applied to the self. In contrast, we suggest that the critical question is how the emotional state of the self is assigned to the other.

In the next section we go on to present the building blocks of the model and briefly sketch how the model architecture may bring about different stages of empathic processing. In the subsequent section we address the 4 questions presented above in the context of the model. Finally, we will discuss how empathy deficits seen in individuals with psychopathy, autism, and alexithymia may be explained by disruption to different aspects of the model architecture. Predictions of the model are detailed throughout the manuscript along with suggestions for further research.
The Self to Other Model of Empathy (SOME)

For clarity, we will start by specifying our working definitions of emotional contagion and empathy. We consider emotional contagion (also known as affective/empathic resonance) to have occurred when exposure to another’s emotion causes the perceiver to be in a matching affective state, but where that affective state is not explicitly recognised as being experienced by the other. For empathy to have occurred, the requirements for emotional contagion have to be met, and in addition the perceiver has to explicitly ‘tag’ their affective state as being experienced by the other.

We suggest that empathy relies on two sources of input derived from computationally intensive processes, which can be considered components of the empathy system itself:

The SITUATION UNDERSTANDING SYSTEM and the AFFECTIVE CUE CLASSIFICATION SYSTEM. The SITUATION UNDERSTANDING SYSTEM is a domain-general appraisal system, which provides an estimate of the emotional state of another based upon the situation they are in. It does not necessarily require the empathiser to have direct experience of the situation, but can provide an estimate of the other’s state as a result of deductive reasoning or associations with relevant stimuli (e.g. black tie = funeral). We suggest that the deductive reasoning process does not necessarily involve ToM, although the ToM System may be recruited in this process. We also suggest that this deductive reasoning process is likely to rely on stored knowledge, with particular emphasis on stored socioemotional knowledge (e.g. funeral = sad). The SITUATION UNDERSTANDING SYSTEM is an input system as it can directly influence the empathiser’s affective state (coded for in the AFFECTIVE REPRESENTATION SYSTEM described below), driving the emotional state of the self towards that of the object of empathy. The influence of the SITUATION
UNDERSTANDING SYSTEM on the AFFECTIVE REPRESENTATION SYSTEM means that this system is sufficient to produce emotional contagion. Meta-analyses such as that of Van Overwalle (2009; see also Barbey, Krueger, & Grafman, 2009) suggest that both dorsal and ventral medial prefrontal cortex are involved in the establishment, storage and use of social script knowledge, while others have noted the involvement of the temporal poles in addition to medial prefrontal cortex in these processes (Frith and Frith, 2003).

The AFFECTIVE CUE CLASSIFICATION SYSTEM performs low-level perceptual categorisation (‘emotional pattern matching’) of person-level cues signalling the affective state of another. These may include facial expression, tone of voice or biological motion cues. A number of regions of inferior and superior temporal cortex have been shown to be involved in processing these cues to another’s emotion (Said et al. 2010; Scott et al., 2009). We argue that this system is sufficient to produce emotional contagion by influencing the state of the AFFECTIVE REPRESENTATION SYSTEM either directly, or via the MIRROR NEURON SYSTEM (see below). The SITUATION UNDERSTANDING SYSTEM and AFFECTIVE CUE CLASSIFICATION SYSTEM can influence each another; person-level cues may affect the appraisal process and vice versa. We characterise these moderating influences of one input system on another with dashed arrows in the model. All later processes rely on the appropriate processing of input.

The MIRROR NEURON SYSTEM, localised to premotor and inferior parietal cortices (Molenberghs et al., 2012), may provide an indirect link between the AFFECTIVE CUE CLASSIFICATION SYSTEM and the AFFECTIVE REPRESENTATION SYSTEM. In general, it consists of associations between the perception and production of actions (Cook et al., 2014). When the actions are, for example, emotional facial expressions,
the MIRROR NEURON SYSTEM may bring about emotional contagion as a result of automatic imitation of the observed emotional expression. Although a matter of debate, several theorists including Charles Darwin (1872) and William James (1890) have argued that production of a particular facial expression may induce the experience of that emotion (see Davis et al., 2010). Therefore, with a limited range of actions, the MIRROR NEURON SYSTEM may lead to emotional contagion but activation of the MIRROR NEURON SYSTEM is not necessary for empathy to occur.

In our model, two representational systems are necessary for empathy. The THEORY OF MIND SYSTEM represents the mental states of the self and others within a propositional system. Several qualitative and quantitative reviews (e.g. Frith and Frith, 2003) have localised the THEORY OF MIND SYSTEM to the temporoparietal junction, medial prefrontal cortex and precuneus. The AFFECTIVE REPRESENTATION SYSTEM, likely localised to the insular cortex and anterior cingulate cortex (Craig, 2009; Critchley et al., 2004), provides a representation of the current affective state of the self. This representation is not necessarily conscious, but is consciously accessible. Following the work of several emotion theorists (e.g Craig, 2009; Bosse, Jonker, & Trure, 2008) we suggest that basic emotional states reflected in the autonomic nervous and sub-cortical emotion circuits are largely unconscious in nature – thus the insula and anterior cingulate cortex provide a (re-) representation of the empathizer’s affective state which is available for introspection. Although always a representation of the empathiser’s own emotion, the representation can be described as a ‘simulation’ of another’s state when the empathiser’s state has been caused by that of the other (via the action of the Self-Other Switch; see next paragraph). Therefore, when Person A is empathic towards Person B, then by definition Person A’s affective state is isomorphic to Person A’s representation of Person B’s affective state.
A component of our model critical for empathy (but not necessarily emotion contagion) is the **SELF-OTHER SWITCH**. The SELF-OTHER SWITCH has two functions. The first is to bias the information processed by the input systems so that one’s own affective state (represented in the AFFECTIVE REPRESENTATION SYSTEM) is appropriate to the other person’s situation/state. Thus, the empathiser’s SITUATION UNDERSTANDING SYSTEM will processes the situation the other is in, rather than the situation the empathiser herself is in. The empathiser will pay greater attention to the other and thus the AFFECTIVE CUE CLASSIFICATION SYSTEM will process the emotional cues displayed by the other. As the ‘feeds’ into the Input Systems change following the application of the SELF-OTHER SWITCH, the resultant affective state of the self will now be appropriate to the judged state of the other. The second function of the SELF-OTHER SWITCH is to tag that the empathiser’s current state is appropriate for the other. This tagging occurs within the ToM system but not within the AFFECTIVE REPRESENTATION SYSTEM. Based on the rich empirical literature on egocentric biases (e.g. Rogers et al., 1977) we suggest the SELF-OTHER SWITCH’s default state is ‘self”, and that switching to ‘other’ is an active process. The neuroanatomical location of the SELF-OTHER SWITCH is more speculative than the other components of the model, but recent work suggests that the temporoparietal junction may be engaged in switching between the processing of self and other representations (Santiesteban et al., 2012; Spengler et al., 2009).

**FIGURE 1 ABOUT HERE**

The difficult mechanistic questions of empathy: Answers from the model
1) How do we know what another is feeling?

A necessary first stage in empathy is the classification of the other’s emotional state. This process is necessarily ‘noisy’ as emotional states are not perfectly predictable from external cues. Therefore, we suggest that in order for empathy to have truly occurred the empathiser’s affective state must be isomorphic to that of another, but that an individual may be said to be empathic if their affective state matches their own classification of the emotional state of another. In other words, true isomorphism defines empathy from a third-person perspective. However, an individual may be empathic from a first-person perspective, yet ‘get it wrong’. To give a crude analogy, a calculator may be working well, but it will not produce the correct answer if the wrong numbers are inputted. In the same way, if any of the input systems used to classify another’s emotional state are dysfunctional, the empathiser will engage in the empathic process, but the resultant affective state will not be isomorphic with the affective state of the object of empathy. We therefore highlight the importance of distinguishing between empathic processing and the outcome of this process.

It is likely that the process of attributing an emotional state to another relies largely upon two sources of information; cues from the person (e.g. facial expression, tone of voice, gait; processed by the AFFECTIVE CUE CLASSIFICATION SYSTEM), and cues derived from the situation that the object of empathy is judged to be in (e.g. Bob is attending a funeral; processed by the SITUATION UNDERSTANDING SYSTEM). Psychological and neurological work over the last two decades (e.g. Johnson and Morton, 1991; Kanwisher et al., 1997) suggests that person-level cues may be processed by specialised perceptual systems supported by dedicated cortical and subcortical neural circuits that instantiate a perceptual ‘pattern-matching’. This pattern-
matching results in a classification of input according to emotion. For example, Winston et al (2004) used functional magnetic resonance imaging to demonstrate that anterior portions of the superior temporal sulcus code for invariant facial features signalling emotional state independently of facial identity. The result of these perceptual classification systems is that low-level classification of emotion is achieved automatically and will feed into the AFFECTIVE REPRESENTATION SYSTEM.

Situational cues allow the attribution of emotion when person-level cues are not present. We suggest that situational cues can contribute to classification of another’s emotion via two routes. The low-level route is addressed here (from the SITUATION UNDERSTANDING SYSTEM to the AFFECTIVE REPRESENTATION SYSTEM), the high-level is addressed in the next section (“What is the role of Theory of Mind in Empathy”). We suggest that emotion can be attributed to another automatically via learned situational cues to emotion. For example, the sight of a black necktie (typical male funeral dress in the UK) may come to serve as an automatic cue to attribute sadness in the wearer. Similarly, wearing a flower in the button-hole of a formal suit (typical male dress for a wedding in the UK) may serve to cause the attribution of joy to another. For both the low-level situational and person-level processes ‘recognition of emotion’ does not imply any overt or conscious recognition of emotion in the other. Rather, they serve as automatic cues to an emotional state.

2) What is the role of Theory of Mind in Empathy?

Under our model ToM and empathy are not identical. While ToM involves the cognitive representation of the mental states of others (beliefs, knowledge, percepts; Frith & Frith, 2003), empathy occurs when classification of another’s emotion causes a corresponding
change in one’s own affective state (emotional contagion) and when one’s own affective state is then recognised as appropriate to another. Thus, empathy and ToM are not necessarily distinguished on the basis of what is represented (e.g. it is possible for both systems to represent fear), but on the basis of the system used to represent the information (i.e. whether the information is represented using a cognitive vs. an affective ‘code’). There is a further subtle distinction between ToM and empathy that we term ‘interpersonal causation’. Interpersonal causation is essential in the case of empathy, but is not required for ToM processing. For example, I can know that you believe in God, I can know what it is to believe in God, but in order for Theory of Mind to be demonstrated there is no requirement that your belief in God has to cause me to believe in God. In contrast, such interpersonal causation is critical in the case of empathy; in order for empathy to be demonstrated recognition of the affective state of another must causally drive the experience of the corresponding state in the self.

So, what role does ToM play in empathy? We suggest that ToM is part of the ‘high-level’ route for processing situational cues and also plays a necessary role in empathy (although it is not necessary for emotional contagion). With regard to situation understanding, there may be times, for example when asked the question “How does Uta feel?” where we use ToM to perform inferences about Uta’s desires and beliefs to infer her affective state. We may have heard that Uta has just achieved a major research prize and so infer that Uta is likely to be very happy. In itself this cognitive process is affectively neutral, however the outcome of this cognitive operation is likely to provoke affective reactions. For example, we may have emotional associations to the word happy, or we may imagine Uta smiling and therefore activate the person-level processes (AFFECTIVE CUE CLASSIFICATION SYSTEM) described above.
The AFFECTIVE CUE CLASSIFICATION SYSTEM, SITUATION UNDERSTANDING SYSTEM, and THEORY OF MIND SYSTEM comprise the inputs into the empathy network and represent the ways in which an affective state is recognised in another. We suggest that the first two systems moderate each other – situation understanding affects how we interpret the person-level cues exhibited by another, and person-level cues affect our situation understanding. It is possible that high-level ToM-based situation understanding is only brought on-line when under volitional control or perhaps when there are large discrepancies between the products of the AFFECTIVE CUE CLASSIFICATION SYSTEM and SITUATION UNDERSTANDING SYSTEM. As these two input systems are proposed to moderate each other, one would expect bi-directional prediction-error signals to be derived when the outputs of the two systems differ. In order to reduce error, the source of the error needs to be identified. If the SITUATION UNDERSTANDING SYSTEM predicts an affective state that is incompatible with the AFFECTIVE CUE CLASSIFICATION SYSTEM then the discrepancy should be used to either 1) update the predictions generated by the SITUATION UNDERSTANDING SYSTEM, or 2) allow for the AFFECTIVE CUE CLASSIFICATION SYSTEM to be recalibrated. Should this error-signal exceed a certain threshold the high-level ToM system may be engaged in order to determine which model should be updated.

Therefore, under the model described above ToM system activity should be predictable based on the degree of discrepancy between the emotional state of another signalled by their situation and the emotion they are portraying. Furthermore, this activity should be linked to an updating of either the SITUATION UNDERSTANDING SYSTEM or AFFECTIVE CUE CLASSIFICATION SYSTEM.

3) How does the state of another cause a corresponding state in the self?
Before discussing the mechanism by which the affective states of another cause corresponding states of the self (i.e. emotional contagion), it is first necessary to discuss representational systems. In contrast to the loose grouping of models commonly referred to as ‘shared network’ or ‘simulationist’ (see Goldman & Sripada, 2005 for discussion), we do not think that the affective state of the other is represented/simulated in one’s own affective system. Rather, we suggest that the affective system is always self-focussed in that the only affective representation possible is the state of the self, perhaps due to the reliance of the AFFECTIVE REPRESENTATION SYSTEM on information relating to the state of the body (e.g. physiological arousal or gustatory state). The mechanism by which we represent the affective state of the other is addressed below (question 4 – How do we come to represent the state of the other?).

We suggest that the three inputs (from the AFFECTIVE CUE CLASSIFICATION SYSTEM, SITUATION UNDERSTANDING SYSTEM, and THEORY OF MIND SYSTEM) to the AFFECTIVE REPRESENTATION SYSTEM all serve to independently bias the AFFECTIVE REPRESENTATION SYSTEM (coding for the affective state of the self) i.e. all inputs change the way the empathiser feels. The process by which the input systems affect the empathiser’s AFFECTIVE REPRESENTATION SYSTEM is not active and/or effortful, rather the product of the input systems automatically drives the empathiser’s affective system (e.g. Seitz et al., 2008; Prochnow et al., 2013). This process is competitive; each input mechanism ‘fights’ for influence on the AFFECTIVE REPRESENTATION SYSTEM (as in the motor system where motor programs compete for control of the final common output path). It is likely that this competitive process is dynamically weighted by standard processes such as attention and motivation (thus allowing for modulation of empathic responses by the context). The competition between inputs may also be weighted in favour of the THEORY OF MIND SYSTEM.
system in order to resolve any conflict arising from discrepant outputs of the

AFFECTIVE CUE CLASSIFICATION SYSTEM and the SITUATION UNDERSTANDING SYSTEM. ¹

One must then ask the crucial question “how is it possible that these input systems bias
the affective state of the self”. We suggest that the mechanism by which it becomes
possible for classification of another’s affective state to influence one’s own affective
state occurs over development. In contrast to a developmental view, several studies
have demonstrated what appear to be innate empathic responses to the emotional
displays of others. The prototypical example of this is the finding that infants as young
as 18 hours old (Martin and Clark, 1982) respond by crying when they hear the cries of
other infants (Sagi and Hoffman 1976; Simner 1971). As defined above this constitutes
emotion contagion (the state of another causes the instantiation of that state in the self),
but there is no evidence that this ‘contagious crying’ constitutes empathy. There is also
no evidence that infants recognise the cries of another infant as an emotional signal at
all – the cries may act as an aversive stimulus like pain or hunger. Hearing another’s
cries may therefore merely act as an ‘innate releasing mechanism’ for the production of
crying. In line with this view there is no evidence of emotional facial expression mirroring
in newborns (Ray and Heyes, 2011), indeed fully differentiated emotional facial
expressions do not emerge until approximately 12 months of age (Bennett et al., 2005;
Geangu et al., 2010; Messinger et al., 1999). The empirical literature on emotional
contagion in neonates therefore does not allow one to draw strong conclusions as to
whether emotional contagion is hard-wired or learned, but this distinction is useful in
classifying developmental theories of emotional contagion.
At the extreme nativist end of the scale are theories such as that forwarded by Meltzoff and Gopnik (1993). These authors argue that infants are innately specified with a primary set of emotions and that these emotions are innately, and bi-directionally, connected to a discrete set of facial expressions. They also argue that infants have an innate ability to imitate (including an ability to imitate facial expression). Accordingly, they posit that observation of another’s emotional facial expression causes the infant to imitate that facial expression which, in-turn, induces the corresponding emotional state in the infant. Finally, Meltzoff and Gopnik suggest that the infant comes to realise that observable human behaviour is caused by internal mental states. Thus, they attribute the emotion to the object of empathy. Although a coherent theory, there is little evidence for a central plank of the Meltzoff and Gopnik theory; that human neonates can imitate facial actions (Anisfeld 1996, 2005; Ray and Heyes, 2011).

In common with the theory of Meltzoff and Gopnik (1993), Preston and de Waal’s (2002) Perception-Action Model of empathy argues that evolution provides the key mechanism by which the emotion of the other can affect one’s own emotion, but also suggests that learning plays some role in empathy. They base their model of empathy on the Ideomotor account of action imitation (Prinz 1997; Greenwald, 1979). Ideomotor theory proposes that perception and action are represented in a ‘common code’, into which dedicated perceptual and motor representations can be translated. It is claimed that the existence of a common code allows perception of an action to automatically activate motor representations of that action. However, the existence of a common code does not solve what Heyes (2001) has termed the ‘correspondence problem’, it merely shifts the locus of the problem. In an affective context the correspondence problem refers to the fact that the emotions of others are encoded perceptually through the ‘distal senses’ – we detect emotions as patterns of light across the retina or a series of sound waves –
and we experience our own emotions as physiological changes with or without accompanying interoceptive awareness. The representation of another’s emotion, and the representation of one’s own emotion are thus in incommensurate codes. What must be explained is how the visual perception of another’s sad face causes the perceiver to feel sad, rather than happy or afraid, for example. Hypothesising an intervening coding system into which perceptual representations of another’s emotion and affective representations of self-emotion can be translated merely moves the correspondence problem to the common representational space; the correspondence problem becomes “how does the perceptual representation of another’s sadness activate the common code for sadness rather than the common code for joy, and how does the common code for sadness activate the affective qualities of sadness in the self rather than a feeling of joy or fear?” (see Brass & Heyes, 2005, for discussion of this problem in the realm of action imitation). According to Preston and de Waal (2002) the answer is that evolution has provided the mappings between perception of another’s emotion and the corresponding common code, and also between the common code and the corresponding emotional state in the self. They argue, however, that experience plays a role in shaping an empathic response, as familiarity with the object of empathy, increased ability to control and regulate emotion, and cultural display rules influence whether an empathic response is observed. Preston and de Waal also allow for experience to ‘fine-tune’ the links between perceived affective states of the other, and corresponding affective states in the self, specifically by contributing to emotion recognition.

A further class of learning-based models argues for a much greater role of experience in the mechanism by which perception of emotion in another causes the corresponding emotion in the self. In these models, experience plays an inductive role (Gottlieb, 1976),
it is necessary for the formation of the mechanism by which perceived emotional states cause activation of matching emotional states in the self. It should be noted that although incompatible with the model forwarded by Preston and de Waal (2002), learning-based models are not necessarily incompatible with ideomotor perception-action models. For example, while representing the adult action imitation system as an ideomotor perception-action model, Elsner and Hommel (2004) provide evidence for the idea that associative learning underlies the formation of the 'common-coding' system.

One such learning-based theory of emotional contagion is that detailed by Heyes and Bird (2007; see also Gergely and Watson, 1996), based on another model of action imitation, the Associative Sequence Learning model (Heyes, 2001; 2010). This theory argues that, at birth, infants lack connections between perceptual representations of emotion in another and their own affective states. That is, (barring the existence of a limited number of innate releasing stimulus links such as hearing crying producing a cry response), perception of another’s emotion will not lead to a matching state in the self. Under this account, emotional contagion occurs via associatively learned bi-directional links between perceptual representations of emotion in another (i.e. representations of emotional facial expressions, bodily postures, and tones of voice) and the corresponding emotional state in the self. These links are largely formed by caregivers imitating the emotional state of the infant via facial expressions and emotionally expressive actions (e.g. clenched fists). The sources of learning by which such links may be formed have been delineated elsewhere (e.g. Heyes, 2001; Allen & Heaton, 2010; Heyes, 2010) and include: direct self-observation (the infant feels happy and laughs, enabling the affective state of being happy to be associated with the sound of a laugh), mirror self-observation (the infant is able to observe their own reflection while producing an emotional facial expression – thus associating the internal affective state
which caused their emotional facial expression to be associated with a perceptual representation of that emotion), synchronous action (both the infant and caregiver respond with fear to a loud noise, enabling the internal fearful emotion to be associated with a perceptual representation of fear), acquired equivalence experience (when both the internal affective state and a perceptual representation of that state in another are both associated with a common third representation, such as a sound like “Yuck” for example), and being imitated by another, typically the primary caregiver (being imitated provides correlated experience of being in an affective state while perceiving the emotional expression corresponding to that state). All of these kinds of experience will result in perception of emotion in another activating in oneself the corresponding affective state via the learned links between perceptual and affective representations of emotion. Of course, individual differences in any of the input systems of the child likely affect the quality of this associative learning process, as would the ability of the caregiver to engage with the child. For example, a fearless temperament in the child may reduce instances of synchronous fear-related action or a depressed caregiver may fail to reliably imitate their child’s positive emotional expressions.

A learning account can explain all three of the routes by which we posit another’s emotion may influence the emotion of the self. First, as already outlined, we suggest that perception of a cue to another’s emotion such as a facial expression or tone of voice can become associated with a corresponding emotion in the self (AFFECTIVE CUE CLASSIFICATION SYSTEM-AFFECTIVE REPRESENTATION SYSTEM Route). Second, the Associative Sequence Learning model, in explaining the development of the ability to imitate and therefore the origin of mirror neurons, explains the mirror neuron mediated link between another’s emotion and a corresponding change in the affective state of the self. Under this model, the bidirectional associative link formed
between the perceptual representation of a person-level emotional cue in another (e.g. an emotional facial expression) and the motor program used to produce the same cue by the self results in automatic imitation of an observed emotional expression by the empathiser (see Wu et al., 2009 for demonstration of the feasibility of this approach). Activation of this motor program, in turn, activates the associated affective state in the empathiser (c.f. facial feedback hypothesis, James, 1890). Third, we suggest that learning plays a necessary role in the high-level route to situation understanding. Earlier we described the use of ToM to solve the problem of determining what emotion another is likely to be experiencing as an affectively neutral cognitive process. However, the outcome of that computational process may result in the individual imagining either the object of empathy experiencing the inferred emotion (and thus activating perceptual representations of emotion as if the object of empathy was in direct observation) or cause activation of an emotional stimulus associated solely with the empathiser’s own affective state. Indeed, a similar mechanism may explain the low-level route to situation understanding if one assumes that salient situational cues (such as the previous examples of black ties at funerals, or the wearing of flowers at a wedding) can enter into associations with affective states. The perception of a situational emotional cue would activate the associated affective state.

Although largely untested in the domain of emotional contagion, a substantial body of evidence supports the Associative Sequence Learning model of action imitation (Heyes, 2010, Cook et al., 2014). Supporting the extension of the Associative Sequence Learning model to emotional contagion are empirical reports demonstrating the opportunity for associations to be learned early in development between internal affective states and perceptual representations of these states in others (see Gergely and Watson, 1996 and Ray and Heyes, 2011, for full details). These include evidence
that: infants find mirror self-observation rewarding and engage in emotional displays to their reflection (Amsterdam, 1972); and, perhaps most importantly, that infants are readily imitated by their caregivers. Infants spend approximately 65% of their waking hours in face-to-face interaction with caregivers (Uzgiris et al., 1989), and an imitative episode occurs approximately once a minute, with the mother imitating the child (rather than the child imitating the mother) 79% of the time (Pawlby, 1977). Crucially for the extension of the Associative Sequence Learning model to emotion contagion, it has been reported that mothers are more likely to imitate their baby’s categorical emotion displays than more ‘random’ facial movements (such as twitches or half smiles) (Malatesta and Izard, 1984; Malatesta et al., 1989). In addition, and somewhat counter-intuitively given the caregiver’s role in their infant’s affective regulation, even infants’ expressions of sadness and anger produce affective responses of sadness and anger in their mothers (Tronick, 1989) and maternal reactions to negative affect include mock expressions of negative affect (Malatesta and Izard, 1984).

Theories such as that of Heyes and Bird (2007) are ‘direct’, i.e. intervening conceptual or higher-order representations are not required to mediate the links between perceptual representations of another’s emotion and an affective reaction in the self. Direct theories specify that should a representation of another’s affective state be activated then (given appropriate experience under learning-based models) there should be a direct and automatic influence of that representation on the empathiser’s own affective state. Note that it is not the case that this influence necessarily results in an isomorphic state in the empathiser (i.e. emotional contagion will not necessarily occur), as the direct route is only one of a number of influences upon the empathiser’s affective state. There are many potential sources of moderation of the effect of another’s emotion on the perceiver’s emotional state such as the empathiser’s resources (trait- or state-based...
emotional, attentional and cognitive resources), the task they are performing, competing
motivation goals, and the relationship between the empathiser and the object of
empathy. Each of these potential moderators may act directly upon 1) the
representation of another’s affective state, 2) the empathiser’s own affective state, or on
the degree to which information related to the other is processed (and therefore the
degree to which person- and situation-level input systems cause activation of an
affective representation), or the degree to which the empathiser’s own affective state
influences behaviour. For example, imagine the situation where a graphic documentary
depicting victims of torture is viewed. The extreme suffering on screen would typically
prompt a highly aversive emotional state in the viewer that they are motivated to reduce.
In response they may pay less attention to what is on the screen, and therefore reduce
the degree to which the onscreen character’s suffering is represented. They may try to
think of happier thoughts, and therefore set-up a competing input into their affective
system, they may apply top-down control to inhibit their own degree of emotional
arousal, and finally, they may maintain an air of indifference in order to control the
expression of their emotion. All of these strategies may be effective but none of them
rely on the direct link between the representation of another’s suffering and a
corresponding state in the self being extinguished.

4) How is the state of the other represented?

We have previously argued that the empathiser’s affective system is not used to
represent the state of the other (the state of the other is represented in the empathiser’s
perceptual and cognitive systems). However, under our working definition empathy is
distinguished from emotional contagion when, in addition to being in an isomorphic state
to the object of empathy, the empathiser recognises that their emotional state is
appropriate for the other. We suggest that the process of ‘tagging’ one’s own emotional state as appropriate for the other is caused by a set of processes that are engaged via the operation of the SELF-OTHER SWITCH.

The SELF-OTHER SWITCH does not bring about a change in the system architecture. Rather, the initial effect of the SELF-OTHER SWITCH is to change the information processed by the three input systems (SITUATION UNDERSTANDING SYSTEM, AFFECTIVE CUE CLASSIFICATION SYSTEM, and THEORY OF MIND SYSTEM) as a result of an attentional change from the default (at least for those who develop in a Western culture; Kitayama and Park, 2010) self-focussed attention, to processing information relating to the other. Thus, person-level cues exhibited by another are more likely to be attended to, and therefore their influence on the AFFECTIVE REPRESENTATION SYSTEM amplified. The SITUATION UNDERSTANDING SYSTEM and THEORY OF MIND SYSTEM will process information relating to the state of the other rather than the self and their output will also influence the AFFECTIVE REPRESENTATION SYSTEM. The combination of these effects will serve to drive the affective state of the self towards that of the other.

The second effect of the SELF-OTHER SWITCH is to create a cognitive representation of one’s own affective state within the THEORY OF MIND SYSTEM, which is ‘tagged’ as belonging to the other and held in working memory. The empathiser’s own affective state may continue to match the cognitive representation of the object of empathy’s affective state (empathy), or it may develop into a different affective state (e.g. sympathy). The act of engaging the SELF-OTHER SWITCH, which changes processing from the default self-focussed to other-focussed state, is an active process. Therefore, switching the SELF-OTHER SWITCH to process the other is likely to be caused either
by a conscious volitional decision, or by a motivational goal / task set which requires
other-focused processing. In the case of empathy the goal may be to gain insight to the
experience of other, which may in turn assist social decision-making.

‘Empathy Disorders’ in the Context of the SOME model of Empathy

The core conceptual implication of our model is that factors affecting one’s own
experience of emotion will determine what emotional associations are learned. Below
we will provide short overviews of three empathy disorders (psychopathy, autism, and
alexithymia), and use the model outlined in this paper to map the source of the empathy
deficit in each disorder, as well as to make some predictions of the model that could be
tested in different populations.

Psychopathy

Psychopathy is perhaps the archetypal empathy disorder. The current clinical definitions
of psychopathy comprise both emotional dysfunction, as well as overt antisocial
behaviour (Hare and Neumann, 2006). The emotional dysfunction dimension includes
reduced empathy and guilt, as well as reduced attachment to significant others. These
symptoms are often considered be the distinctive feature of psychopathy (Blair et al.,
2005; Viding & Frick, 2009), and the ability of psychopaths to inflict serious harm on
others suggests a profound disturbance in the appropriate ‘empathic’ response to the
distress of another.

In addition to a lack of empathy towards others, reduced or atypical experience of one’s
own emotion has long been considered a hallmark of psychopathy (Cleckley 1941) –
with empirical studies demonstrating qualitative and/or quantitative differences in emotional experience (Brook et al., 2013; Hare, 1993, 1998; Steuerwald and Kosson, 2000) and the use of affective language (Gillstrom and Hare, 1988; Hare and McPherson, 1984; Williamson et al., 1991). Other markers of emotional experience have also been found to be absent or reduced in psychopathy, such as physiological responses to affective material (Hare, 1978; Patrick, et al., 1993) and memory for emotional events (Christianson et al., 1996). Two recent studies indicate that children and adolescents with psychopathic tendencies report lower fear reactivity/arousal (compared with typically developing peers or peers with autism or conduct problems) when they imagine themselves in scary and dangerous situations (Jones et al., 2010; Marsh et al., 2011).

In common with other researchers (e.g. Blair et al., 2005) we suggest that these data indicate that individuals with psychopathy have an atypical experience of, at least, fear and sadness (but possibly of other emotions as well, see Brooks et al., 2013). Under our model, a deficit in e.g. the experience of distress emotions would lead to a diminished ability to recognize distress cues in others and therefore a lack of empathy for distress. We have argued that emotional contagion, which is necessary for empathy to occur, develops through repeated pairing of an affective state with cues to that state in another. If the infant experiences distress less often than is typical, there will be reduced opportunity for the infant to learn which cues reliably signal distress in another. Reduced distress in the infant would thus result in fewer opportunities for learning and mean that individuals with psychopathy would lack an experiential understanding of the distress state - leading to a lack of emotional contagion and empathy for distress emotions. Recent data has also emerged suggesting that children with psychopathic tendencies do not seek out eye-contact with their mothers (although the mothers
themselves do not differ from mothers of typically developing children in seeking out eye contact with their children, e.g. Dadds et al., 2014). It is unclear what drives the reduced drive to seek eye contact in these children, but this is likely to further reduce learning opportunities about emotions over development. Whereas non-psychopathic individuals would typically empathise with another’s distress and therefore desist in carrying out behavior evoking these emotions in another, those with psychopathy would not have learned the necessary links supporting emotional contagion that facilitates empathic processing and appropriate behaviors in the face of another’s distress.

In support of this suggestion are findings of a reduced autonomic response in children and adults with psychopathic traits to stimuli portraying fear and sadness in others (Aniskiewicz, 1979; Blair, 1999; Blair et al., 1997; de Wied, van Boxtel, Matthys, & Meeus, 2012; House and Milligan, 1976). Atypically low amygdala and insula activity to other people’s distress and pain has also been documented in individuals with psychopathy/psychopathic tendencies during both prolonged viewing conditions, as well as pre-attentive stimulus presentation (e.g. Dolan and Fullam, 2009; Marsh et al., 2008; Jones et al., 2009, Lockwood et al., 2013; Sebastian et al., 2012; Viding et al., 2012).

Further support is provided by studies of emotion recognition, which suggest that psychopaths have difficulties in recognising fearful and sad facial and vocal expressions (See e.g. Blair et al., 2001; Blair et al., 2005; Dolan and Fullam, 2006; Fairchild et al., 2009; Kosson et al., 2002; Stevens et al., 2001; although see Hastings et al., 2008 for conflicting findings) and show diminished processing of fear and disgust at a pre-attentive level (Sylvers, Brennan, & Lilienfeld, 2011). Children with psychopathic tendencies also self-report lower levels of emotional contagion and empathy (compared with typically developing, ASC, and conduct problem peers) in response to vignette
stories and film clips of other people experiencing emotions, particularly fear and sadness (e.g. de Wied et al., 2012; Jones et al., 2010; Pardini, et al., 2003; Schwenck et al., 2012).

Despite the apparent deficits in emotional contagion, emotion processing and empathy for distress, adults with psychopathy and children with psychopathic tendencies are not impaired in all aspects of understanding others. They are characterized by their ability to manipulate others (Hare, 2003), an ability which requires good ToM skills. Several studies employing experimental ToM paradigms have reported intact ToM in individuals with psychopathy / psychopathic tendencies (Dolan and Fullam, 2004; Jones et al., 2010; Richell et al., 2003; Schwenck et al., 2012).

With reference to the model architecture, individuals with psychopathy may have a conceptual representation (THEORY OF MIND SYSTEM) of other people’s distress and may be able to correctly label distress emotions (even if this process is more effortful and error prone for them), whilst lacking the appropriate arousal response to another person’s distress. The model predicts that typically developing individuals will utilise the THEORY OF MIND SYSTEM when SITUATION UNDERSTANDING SYSTEM and AFFECTIVE CUE CLASSIFICATION SYSTEM inputs are conflicting (i.e. when someone looks upset despite being in a situation expected to provoke joy). In contrast, we would expect that individuals with psychopathy would not recruit their THEORY OF MIND SYSTEM under these conditions, as their impoverished AFFECTIVE CUE CLASSIFICATION SYSTEM would provide only weak or no input and therefore a there would be a smaller likelihood of a conflict arising between the two input systems.

We have argued that psychopathy is primarily a disorder characterised by atypical affective experience, which in turn restricts the situations in which individuals with
psychopathy can feel empathy with other people. The basic premise of the model
detailed above is that one’s own emotional state is attributed to another when
empathising, it therefore follows that one’s own emotional experiences restrict the kind
and degree of empathy that can be felt.

Autism

Autism is a developmental disorder characterised by impaired social and
communication skills and a restricted repertoire of interests and activities (APA, 2013).
Individuals with autism have frequently been characterised as lacking in empathy (e.g.
Baron-Cohen and Wheelwright, 2004; Gillberg, 1992). In support of this
characterization, individuals with autism score lower on the Empathy Quotient (Baron-
Cohen and Wheelwright, 2004) which assesses the self-reported capacity to take
another person’s mental perspective and the capacity to share their feelings. The claim
of a global empathy deficit in autism does not always reflect the distinction between
emotion recognition, ToM and empathy however. For example, a test widely used in
autism research as a marker for empathy is the ‘reading the mind in the eyes test’
(Baron-Cohen et al., 1996; Baron-Cohen et al., 1997; Baron-Cohen et al., 2001). This
test requires one to recognise the internal state from the eye region of facial
expressions, but does not directly measure the emotional response elicited by the
expression. Under our working definition, therefore, this test does not index empathy.

Several decades of research confirms that autism is associated with a ToM deficit (see
Frith & Happé, 2005, for a review). Under our model, a complete lack of ToM would be
expected to result in a complete lack of empathy, as ToM is necessary for the process
by which one’s own affective state is tagged as relevant to the other. In practice, it
seems probable that this process makes little demand upon the ToM system, and so
empathy is likely to be intact in all but the most severe cases of ToM impairment.

However, even a moderate degree of ToM impairment is likely to impact upon empathy when ToM is required in order to classify another’s affective state i.e. in situations where ToM provides an input into the empathy system. In this case ToM impairments may lead to misclassification of another’s affective state, leading either to an incorrect, non-isomorphic empathic response, or to an absence of an empathic response. A similarly absent, or incorrect, empathic response may arise from atypical social scripts due to reduced attention to, or opportunity for, social contact in autism, impacting upon the SITUATION UNDERSTANDING SYSTEM. As already mentioned, the model predicts that typically developing individuals will utilise the THEORY OF MIND SYSTEM when SITUATION UNDERSTANDING SYSTEM and AFFECTIVE CUE CLASSIFICATION SYSTEM inputs are conflicting. In common with individuals with psychopathy, we would expect that for individuals with autism there would be a reduced probability of a conflict arising between the two input systems (if reduced social attention and / or motivation results in impairments in these systems), and, in addition, the THEORY OF MIND SYSTEM would be less effective in arbitrating between these systems due to ToM impairments in individuals with autism.

Impairments in the SITUATION UNDERSTANDING SYSTEM and THEORY OF MIND SYSTEM may result in incorrect input into the empathy system, and therefore lead to impaired empathy, but we suggest that should another’s emotional state be correctly classified individuals with autism are likely to exhibit intact empathy (providing any ToM impairment is not too severe). This is because the mechanisms by which another’s affective state cause a corresponding change in one’s own affective state are intact in autism: although there are some conflicting reports, the consensus is that the mirror neuron system functions typically in autism (e.g. Hamilton, Brindley & Frith, 2007; Cook...
and Bird, 2012), and associative learning is intact in individuals with autism (e.g., Boucher and Warrington, 1976; Williams et al., 2006). Furthermore, the AFFECTIVE REPRESENTATION SYSTEM is likely to be intact in those with autism, at least as indexed by typical response in the anterior insula when introspecting on emotion and the consistency of emotional self-report (Silani et al., 2008; Berthoz and Hill, 2005). It should be noted, however, that autism may be characterised by reduced social motivation and attention (e.g. Dawson, Webb, & McPartland, 2005; Jones, Carr & Klin, 2008; but see Johnson, 2014; Shah et al., 2013 for an alternate view). Therefore individuals with autism may be less likely to seek-out, and/or attend to social interactions. In turn, this reduced social interaction would result in less opportunity to learn the links between experienced and perceived emotional states and may result in a delay in the development of emotional contagion and empathy.

It is interesting to speculate whether autism is also associated with a SELF-OTHER SWITCH that is overly biased towards the self. As noted above, one of the functions of the SELF-OTHER SWITCH is to direct attention towards others and individuals with autism exhibit less social attention. Indeed, Kanner (1943) described children with autism as displaying ‘extreme egocentrism’. If the SELF-OTHER SWITCH was functioning abnormally in autism one may also expect that, when others are attended, that those with autism would be less likely to draw a sharp distinction between the self and the other and so may be more affected by another’s state than typical individuals. Some evidence for this hypothesis can be found in the action domain, where individuals with autism demonstrate echolalia and echopraxia clinically, and also exhibit a tendency to ‘hyper-imitate’ in lab-based experiments (Spengler, Bird & Brass, 2010). Evidence also exists in the affective domain, where individuals with autism often report greater
personal distress than typical individuals upon perceiving the distress of another (Rogers et al., 2007, see Smith 2009 for an overview). If SELF-OTHER SWITCH function is impaired in autism then one would expect the increased personal distress (due to greater emotional contagion) to be exacerbated by the fact that those with autism may be less likely to form a cognitive representation of their affective state and assign it to the other. It has previously been argued that forming such a cognitive representation is likely to serve to reduce empathic personal distress (Decety and Lamm, 2006), and therefore one might predict that not only would self-reported personal distress scores be higher than in typical controls, but also that the state of high personal distress would be longer-lasting than in typical individuals.

Locating any empathy deficit in individuals with autism within the model described above is likely to have profound implications concerning interventions. For example, if an individual with autism is completely unable to represent other humans as having mental states (which is extremely rare), it is likely that intervention in any form will be unsuccessful in producing an empathic response. However, a complete lack of ToM does not preclude the development of emotional contagion, and, in turn, appropriate responses to another's affective state. The model suggests that interventions should take the form of training in which the experience of emotion is linked to the expression of that emotion in others in order to increase the likelihood of emotional contagion. In addition, social skills training should focus on the development of socio-emotional scripts to improve the accuracy of the SITUATION UNDERSTANDING SYSTEM, leading to higher-quality input into the empathy system.

The empathy impairments attributed to autism are relatively mild under our model, and
depend largely upon the degree of ToM impairment in any one individual with autism.

This picture is at odds with the commonly-held perception that autism is associated with severe empathy deficits. We suggest there may be two reasons for this. The first is definitional; as already described ToM and empathy are fractionated under our model while under alternative accounts the distinction between ToM and empathy is less clear-cut. The second reason may be the high degree of co-morbidity between autism and alexithymia, which we turn to next.

**Alexithymia**

Alexithymia is a subclinical phenomenon marked by difficulties in identifying and describing feelings, difficulties in distinguishing feelings from the bodily sensations of emotional arousal, and a pre-occupation with externally-oriented thinking (Nemiah et al., 1976). Typically measured by self-report, highly alexithymic individuals report agreement with statements such as “When I am upset, I don’t know if I am sad, frightened or angry.” Alexithymia is thought to characterize 10% of the general population (Linden et al., 1995; Salminen et al., 1999), and has been associated with a number of conditions including eating disorders, depression, schizophrenia and substance abuse (Bird & Cook, 2013). A strong association with autism has been demonstrated. Although alexithymia is neither necessary nor sufficient for an autism diagnosis, recent studies have found severe degrees of alexithymia in approximately 50% of individuals with autism, with the majority showing at least some degree of alexithymia (Hill et al., 2004; Lombardo et al., 2007; Nemiah et al., 1976, Bird and Cook, 2013).

We suggest that the primary impairment in Alexithymia lies within the AFFECTIVE REPRESENTATION SYSTEM, and that this impairment is likely to impact upon
empathy processing in a number of ways. An impaired AFFECTIVE REPRESENTATION SYSTEM is consistent with the diagnostic characterisation of alexithymia as a disorder in which individuals are aware that they are having an emotion, but are unsure as to what emotion they are experiencing. If the AFFECTIVE REPRESENTATION SYSTEM is damaged alexithymic individuals will not be able to form a consciously-accessible representation of their own affective state. Under the learning model described above, a lack of differentiated affective states in the self provides no opportunity for affective internal states to be associated with perceptual cues to those states in others.\(^2\)

Co-occurring alexithymia may explain a number of the affective impairments seen in autism. First, evidence for an impairment in the AFFECTIVE REPRESENTATION SYSTEM in alexithymic individuals was provided by Silani et al. (2008). This study showed that impairments in the ability to represent and report one’s own emotion was associated with alexithymia, and that this ability was correlated with activity in anterior insula. Furthermore, levels of alexithymia and anterior insula activity were also associated with self-reported empathy scores. The relationship between alexithymia, empathy, and anterior insula activity was equivalent for both individuals with autism and typical adults, suggesting that previous reports of impaired emotional introspection in autism was, in fact, due to the high proportion of alexithymic individuals within the population of individuals with autism.

This conclusion was supported by the results of a second neuroimaging study, which examined empathic anterior insula response to the pain of another (Bird et al., 2010). In this study participants received a painful electric shock to their own hand in order to map out those areas of the brain which responded when the participant experienced pain.
(their ‘pain matrix’). They then observed abstract cues (coloured arrows), which informed them that a loved one was experiencing a similar shock, and empathic brain activity was measured (operationalized as neural activity in response to the pain of another in one’s own pain matrix). Results showed that the amount of empathic brain activity was predicted by the degree of alexithymia in both those with autism and those without autism, and that the presence of autism was not associated with a reduced empathic brain response.

The imaging studies described above support the hypothesis of an impaired AFFECTIVE REPRESENTATION SYSTEM in alexithymic individuals, and that this impairment results in a lack of empathy. Importantly, empathic responding in these studies did not rely on accurate perception of affective cues displayed by another – therefore the reduced empathy displayed by alexithymic individuals in these studies cannot be explained by perceptual deficits.

When detailing the model it was argued that an atypical AFFECTIVE REPRESENTATION SYSTEM is likely to result in an impaired ability to recognise emotions as a lack of differentiated affective states will result in reduced opportunities to learn the cues to those states in another. Therefore, although the findings detailed above cannot be due to a deficit in emotion recognition, deficits in emotion recognition in alexithymia are predicted by the model. This prediction was recently supported by Cook et al. (2013). Alexithymic and non-alexithymic individuals, with and without comorbid autism, were asked to do two tasks. The first task required participants to decide whether two faces were the same, or different. The faces varied either in identity or emotional expression. In order to perform well on this task participants need to be able to form an accurate percept of the emotion or the identity depicted by the face stimuli,
but do not need to be able to recognise them. Both the alexithymic and autistic participants were unimpaired on this task. The second task used the same stimuli, but required participants to recognise the displayed emotion or identity. Again, the autistic individuals without comorbid alexithymia were accurate on both the identity and emotional expression version of the task but alexithymic individuals, irrespective of whether they had accompanying autism, were unable to recognise the emotional expressions even though they were able to recognise the identities. This result is striking as the alexithymic participants were able to correctly differentiate expressions of anger and disgust, for example, but were unable to label one type of expression anger and one disgust. Such a pattern of results would be expected if the AFFECTIVE REPRESENTATION SYSTEM was not providing a conscious representation of affective states, but the AFFECTIVE CUE CLASSIFICATION SYSTEM was functioning correctly.

The studies investigating the contribution of alexithymia to emotional impairments normally attributed to autism support a series of studies carried out with alexithymic individuals with other comorbid disorders (typically eating disorders). These studies have demonstrated impaired recognition of emotion in alexithymia using face (Jessimer and Markham, 1997; Lane et al., 1996; Lane et al., 2000 Mann et al., 1994; Parker et al., 1993; Prkachin et al., 2009 (although see Kessler et al., 2006 for discrepant findings), and also verbal stimuli (Lane et al., 1996). Paralleling the findings of reduced empathic brain response to the pain of others (Bird et al., 2010), Moriguchi et al. (2007) showed participants photographs of hands and feet in painful situations. Alexithymic individuals showed reduced activity in areas previously reported to be part of the pain matrix, including dorsal ACC. The authors suggest this is indicative of reduced empathy for pain in the Alexithymia group, however, the alexithymia group showed increased activation in other parts of the pain matrix including ventral ACC and anterior insula. The
conflicting results between the Moriguchi study and that of Bird and colleagues are difficult to interpret as stimuli in the Moriguchi were photographed from the first-person perspective (and therefore participants may have imagined their own body in the painful situation) and also as the Alexithymia group rated the images as less painful than the comparison group. As the Alexithymia group rated the images as less painful then even if their empathy system was as effective as the comparison group, they would have experienced less empathy as a result. It should also be noted that, under our working definition of empathy, the neurological activity seen in the Moriguchi and Bird studies may be an index of emotional contagion rather than empathy. It is clear then, that the neurological response to another’s pain requires further investigation in individuals with alexithymia.

Summary

We have outlined a model of the necessary system architecture and information processing needed for empathy to occur. We aimed to provide such a model in order to locate potential impairments within the empathy system which may explain conditions in which empathy is absent or reduced. The model may also provide a framework within which neuroimaging work on empathy can be interpreted.

We used this model to answer four questions that we consider to be crucial in understanding the mechanisms which give rise to empathy. We suggested that domain-general (SITUATION UNDERSTANDING SYSTEM and THEORY OF MIND SYSTEM) and domain-specific (AFFECTIVE CUE CLASSIFICATION SYSTEM) input mechanisms classify the affective state of another. Emotional contagion arises over development via associative learning – cues to the affective state of another become associated with
corresponding affective states in the self. These learned links result in the empathiser’s affective state being automatically driven towards that of the other. Emotional contagion ‘becomes’ empathy via the activation of a self-other switch. This switch causes the inputs into the AFFECTIVE REPRESENTATION SYSTEM to process information relevant to the other rather than the self, and causes the resultant affective state experienced by the self to be stored as a cognitive representation within the THEORY OF MIND SYSTEM and to be tagged as belonging to the other.

This framework was used to explain the empathy deficits observed in psychopathy, autism and alexithymia. It was suggested that psychopathy results in a selective impairment of the experience of fear and sadness, with a resultant failure to associate those affective states with cues to their instantiation in others. In contrast, any empathy impairment in autism was suggested to arise from ToM problems, but only when those impairments are severe. Empathy impairments in Alexithymia were suggested to arise from a more global (in comparison with psychopathy) impairment in the ability to represent one’s own emotion, resulting in an inability to learn associations between cues to affective states in others, and affective states in the self.
Notes

1. It should be recognised that there is an obvious disconnect in the model. We describe input systems (SITUATION UNDERSTANDING SYSTEM & AFFECTIVE CUE CLASSIFICATION SYSTEM) as being directly connected to the AFFECTIVE REPRESENTATION SYSTEM, which provides a representation of the empathizer’s affective state. In reality inputs are likely to influence the empathizer’s affective state itself – and therefore affect the autonomic nervous system and sub-cortical emotion circuits - which then feed-forward information to the insula / anterior cingulate cortex which together comprise the AFFECTIVE REPRESENTATION SYSTEM. We have avoided detailing this step of the model as it is still a matter of great theoretical debate (see discussion in Seth, 2013; Bosse, Jonker, & Trure, 2008). More importantly, given the requirement for conscious experience of emotion under our working definition of empathy, whatever mechanism is responsible for the conscious experience of emotion will, by definition, impact upon the AFFECTIVE REPRESENTATION SYSTEM.

2. The issue of whether Alexithymia provides an opportunity for affective states to be associated with cues to those states in others is relevant to an on-going debate within the field of alexithymia research concerning whether Alexithymia is a unitary phenomenon (e.g. Bagby et al., 2009) or whether two types can be distinguished (e.g. Bermond et al., 2007). Those favouring the existence of two types of alexithymia suggest that Type 1 alexithymia is characterised by blunted emotional response at the physiological level and a reduced ability to introspect upon emotions, while individuals with Type 2 alexithymia have a typical physiological emotional response but a reduced ability to gain conscious access to their emotional states. While both types of Alexithymia would be characterised by an impaired AFFECTIVE REPRESENTATION
SYSTEM - as evidenced by their reduced ability to consciously represent their emotional state - those with Type 2 alexithymia would have typical affective physiological states which may enter into associations with cues to those states in others. It is possible that forming these low-level associations may therefore be a useful therapeutic approach for those with Type 2 alexithymia (Allen & Heaton, 2010), but would be unavailable for those with Type 1 alexithymia.
FIGURE CAPTIONS

Figure 1 – The box represents the input and representational systems involved in understanding one’s own emotion. Note that while a representational system, the Theory of Mind System can also directly influence the Affective Representation System and therefore act as an input system. Empathy occurs via the action of the Self/Other Switch which serves to bias input into the system so that it is appropriate for the other and assigns the resultant feeling state to the other. Emotion contagion, in contrast, can occur without the Self/Other switch and does not require the Theory of Mind system.
References:


http://dx.doi.org/10.1177/0956797612463582


http://dx.doi.org/10.1038/nn1176


James, W., 1890. The principles of psychology. 2 (1890). MacMillan.


51


