EMPATHY IS NOT IN OUR GENES

Cecilia Heyes

All Souls College & Department of Experimental Psychology

University of Oxford

Oxford OX1 4AL

United Kingdom

cecilia.heyes@all-souls.ox.ac.uk

Abstract

In academic and public life empathy is seen as a fundamental force of morality – a psychological phenomenon, rooted in biology, with profound effects in law, policy, and international relations. But the roots of empathy are not as firm as we like to think. The matching mechanism that distinguishes empathy from compassion, envy, schadenfreude, and sadism is a product of learning. Here I present a dual system model that distinguishes Empathy^{1}, an automatic process that catches the feelings of others, from Empathy^{2}, controlled processes that interpret those feelings. Research with animals, infants, adults and robots suggests that the mechanism of Empathy^{1}, emotional contagion, is constructed in the course of development through social interaction. Learned Matching implies that empathy is both agile and fragile. It can be enhanced and redirected by novel experience, and broken by social change.

Keywords

Affect mirroring; affective empathy; associative learning; emotional contagion; empathy; empathic understanding; Learned Matching; mirror neurons; self-stimulation; synchronous emotion.

Introduction

Some people think that empathy, feeling what others are feeling, is a wonderful thing – a friend of social justice, good parenting, humane healthcare, and life-enhancing personal relationships (Batson, 2011; Zaki, in press). Others believe that empathy makes us innumerate and biased – inclined to favour the few over the many, members of our own clique over deserving strangers (Bloom, 2017). Opinions about the value of empathy are deeply divided, but many researchers, politicians and social commentators on both sides of the debate assume that humans are born with a propensity to feel what others are feeling, an instinct favoured by evolution to make us better parents and players for the local team (Bazelgette, 2017; de Waal and Preston, 2017; Preston and de Waal, 2002). This article challenges that assumption.

What is empathy?

Concepts like **empathy** (see Figure 1), that are important in both science and everyday life, attract a variety of definitions (Batson, 2009). Scientists with a special interest in how we understand the minds of others tend to define empathy, or 'cognitive empathy', as a component of **mindreading** (Ickes, 1997). On this view, I empathise with you whenever I understand that you are sad. It does not matter whether knowing this about you makes me sad, or happy, or has no emotional effects on me at all. On the other hand, those who regard empathy primarily as a force for good in public and personal life tend to identify empathy with any sort of generous or compassionate feeling towards others (Batson et al., 2005; Pavey et al., 2005). My wish for you to do well, and to avoid suffering, is empathic regardless of where it came from and what you are currently feeling. My emotional state need not be similar to yours.

However, most contemporary neuroscientists, psychologists and philosophers anchor their definition of empathy to a matching relation between the emotions of two people, an 'agent' and a 'target'. An agent's response to a target's emotion is empathic if it is caused by and resembles, or matches, the target's emotion. I am empathising if your sadness makes me feel sad, but not if it brings me pleasure. Most cognitive scientists also agree that the matching relation that defines

empathy can be produced in a simple way, known as 'emotional contagion' or 'experience sharing', or in a more complex way, known as 'empathic understanding', 'affective empathy' or 'emotional empathy' (Decety and Meyer, 2008; Zaki, 2014).

Building on these areas of agreement, the dual system model in Figure 2 proposes that empathic responses can be produced by either or both of two functional systems. The first system, Empathy^1, operates **automatically**, develops early in humans, and is found in a wide range of other animals. The second, Empathy^2, involves **controlled processing**, develops later, and, insofar as the controlled processing involves mindreading, may be uniquely human. Empathic responses produced solely by Empathy^1 are usually described as 'emotional contagion', whereas those produced solely by Empathy^2, or by the combined operations of Empathy^1 and Empathy^2, provide examples of 'empathic understanding'.

Empathy begins with an emotional stimulus relating to another agent, the 'target' (see left side of Figure 2). This may be a facial or bodily gesture (e.g. wincing or punching the air); a vocalisation (screaming or laughter); an emotive situation (a needle entering flesh or a person winning a race); or a verbal description evoking imagination of such events. In Empathy^1, the emotional stimulus automatically triggers a motoric and/or somatic response via neural circuits in areas including the premotor cortex, inferior parietal lobe, and posterior superior temporal sulcus (motor activation), and the anterior insula and anterior cingulate cortex (somatic activation) (Zaki and Ochsner, 2012). This activation is 'automatic' in the sense that it is rapid, makes little demand on executive function, and is minimally dependent on the agent's intentions. The automatic response - which may or may not be detectable by a casual observer or even by the agent herself - consists of behavioural and/or visceral changes that make the emotional state of the agent more like that of the target. For example, the sight of frowning triggers a frowning response in the agent's facial muscles, and the sight of a needle entering flesh triggers fear-related elevation of heart rate and respiration.

At least in humans, the automatic response that is the output of Empathy^1 often becomes an input for Empathy^2 (see right side of Figure 2). Controlled processes, mediating goal-directed action, appraise the automatic response alongside information about the context of the emotional stimulus

(e.g. fictional or real), the agent's relationship with the target (e.g. in-group or out-group), and the agent's current priorities (Wondra and Ellsworth, 2015, Zaki, 2014). More broadly, Empathy^2 processes information about the outcomes of various actions taken in the presence of this kind of emotional stimulus in the past.

The controlled processing of Empathy^2 may involve metacognitive as well as cognitive appraisal. For example, the agent may conceptualise the target as being in a particular emotional state ('He is anxious'), conceptualise herself as being in the same emotional state ('I am anxious'), and infer that the target's state caused her own state ('I am anxious because he is anxious') (De Vignemont and Singer, 2006).

Both cognitive and metacognitive appraisal can have two kinds of effect. They can modulate the output of Empathy^1, amplifying or dampening the automatic response, and they can produce a controlled response, an action involving approach to, or avoidance of, the emotional stimulus. Whether they result from metacognitive appraisal, or just cognitive appraisal, controlled responses may be prosocial or antisocial – they may help and comfort, or thwart and disturb, the target.

In principle, Empathy^2 could operate alone (dashed lines in Figure 2). Cognitive and **metacognitive** appraisal could yield a matching emotional state without automatic motoric and/or somatic activation (Empathy^1). How commonly this occurs, in the laboratory and in everyday life, is an open empirical question. Several considerations suggest that Empathy^2 rarely acts alone:

1) Studies using behavioural and physiological measures suggest that automatic matching responses to emotional stimuli are ubiquitous in adult humans, leaving little opportunity for Empathy^2 to function alone. For example, electromyographic (EMG) recording from muscles in the face and arms, and measurement of changes in heart rate, respiration, and sweating, have indicated rapid, unintended matching responses to facial expressions of happiness, anger and fear (Dimberg, 1982; Kelly et al., 2016; Moody et al., 2017); body postures associated with anger (Berger and Hadley, 1975); and non-verbal vocalisations expressing happiness, sadness, anger and disgust (Hawk et al., 2012).

2) Meta-analysis of functional magnetic resonance imaging (fMRI) data indicates that certain areas of the brain are active during the generation of all empathic responses, regardless of whether the task encourages or discourages appraisal (Empathy^2), and many of these common areas – such as the supplementary motor area, anterior insular and anterior cingulate cortex - are associated with automatic motoric and somatic activation (Empathy^1) (Fan et al., 2011; Gonzalez-Liencres et al., 2013; Lamm et al., 2011).

3) Research using event-related potentials (ERPs) to study empathic pain has found an early (~140ms) neural response, indicative of automatic activation (Empathy^1), which occurs both when subjects are distracted, and when they are encouraged to appraise the emotional stimulus (Empathy^2) (Fan and Han, 2008; Decety and Cowell, 2014).

These three lines of evidence suggest that Empathy^1, automatic activation of matching motoric and somatic responses, plays a crucial role in human empathy. The cognitive and metacognitive processes of Empathy^2 are needed to select and launch intentional empathy-based action – prosocial or otherwise – but Empathy^1 is the matching mechanism that makes *my* body feel *your* emotion.

Where does Empathy^1 come from?

It is widely assumed that Empathy^1 is an innate mechanism; that we humans, and some other animals, have a genetically inherited, developmentally **canalized** propensity to respond automatically to emotional stimuli with matching emotion (de Waal and Preston, 2017; Doherty, 1997; Gonzalez-Liencres et al., 2013; Hatfield et al., 2014; Hoffman, 2001; Meltzoff, 2011; Pavey et al., 2012; Preston and de Waal, 2002). This nativist view is especially prominent in Preston and de Waal's 'perception-action model' of empathy (PAM; de Waal and Preston, 2017; Preston and de Waal, 2002). The nativist view suggests that during the early evolution of mammals, when parental care was becoming important, and during primate evolution, when cooperation among group

members was increasingly at a premium, natural selection favoured genes promoting Empathy[1]. For example, individuals who had a propensity to respond to distress with distress, rather than aggression, were more likely to survive and reproduce because they were better able to care for their offspring, and more likely to be alerted to danger by others in their social group.

The nativist view is plausible because emotional contagion has been found in a range of animal species, and we are apt to assume that rapid, unintentional, adaptive responses – especially those we share with other animals - are due to dedicated, genetically inherited mechanisms (de Waal and Preston, 2017; Preston and de Waal, 2002). However, the evidence to be surveyed in this article suggests that Empathy[1] is constructed in the course of development by domain-general processes of associative learning. According to this **Learned Matching** hypothesis, the automatic activation characteristic of Empathy[1] depends on a set of learned associations. Each of these **matching emotional associations** connects, in a bidirectional excitatory way, a distal sensory cue (e.g. an emotional facial gesture or vocalisation) with a motoric or somatic response belonging to the same emotional category (Bird and Viding, 2014; Heyes and Bird, 2007). These connections are forged in situations where the experience of a particular emotion 'from the inside' is correlated with observation of the same emotion 'from the outside'.

In the case of vocalisations, such as crying, and some bodily expressions of emotion, such as fist clenching, the association-building situations do not necessarily involve another agent - self-stimulation is sufficient. When a baby hears herself crying she is exposed to a contingency or correlation between the sound of crying (a distal sensory cue) and feelings of distress. The sound of hunger, fear or pain 'on the outside' predicts, and is predicted by, the feeling of hunger, fear or pain 'on the inside'. However, many of the associations mediating Empathy[1] require social interaction for their development. I cannot see my own facial expressions (e.g. wincing) or whole-body movements (e.g. punching the air) as they appear when produced by others; facial expressions and whole-body movements are perceptually opaque (Heyes and Ray, 2000; Ray and Heyes, 2011). Matching emotional associations for perceptually opaque cues are produced by **synchronous emotion** and **affect mirroring** (Papousek and Papousek, 1987). Synchronous emotion occurs when two or more

agents react emotionally to an event in the same way at the same time. In a crowd at a soccer match, I see other fans wincing when I am disappointed, scowling when I am angry, and punching the air when I am feeling elated. Affect mirroring, which has long been known to play important roles in the differentiation of emotional states and affect regulation in infancy (Gergely and Watson, 1996; Parsons et al., 2017), occurs when caregivers imitate infants' facial and vocal expressions of emotion, reacting to joy with joy, surprise with surprise, and even to negative emotions, anger and sadness, with matching affective displays (Malatsta and Izard, 1984; Malatesta et al., 1989; Tronick, 1989).

It has been suggested that mirror neurons implement a range of psychological functions including empathy (Christov-Moore et al., 2014; Gallese et al., 2004; Rizzolatti and Caruana, 2017). To the extent that mirror neurons implement Empathy^{1}, the Learned Matching hypothesis is consistent with theories that assign an important role to learning in the development of mirror neurons (Cook et al., 2014; Giudice et al., 2009; Heyes, 2001; Keysers and Gazzola, 2014; Keysers and Perrett, 2004; Kilner et al., 2007).

Learned Matching does not imply that learning is solely responsible for the development of empathy in general, or Empathy^{1} in particular. The development of every biological characteristic depends on a rich, turbulent stew of genetic and environmental factors (Heyes, 2018). There are no pure cases of nature or of nurture, and there is clear evidence of a genetic contribution to individual differences in empathy (Flom and Saudino, 2017; Uzefovsky et al., 2015; Warrier et al., 2018). However, to date, behavioural genetics has not shown that heritable genetic factors contribute to the matching characteristic of Empathy^{1}; it has not shown that genetic factors promote the development of a mechanism producing matching, rather than nonmatching, emotional responses to emotional stimuli. Twin studies use behavioural measures that conflate emotion understanding (in which there is no emotional matching), Empathy^{2}, and Empathy^{1}. Therefore, it is possible that the genetically inherited psychological endophenotype is not a matching mechanism, but processes involved in social motivation, emotion identification, or emotion regulation (Coll et al., 2017; Decety

et al., 2018; Quattrocki and Friston, 2014). Evidence implicating genes related to the oxytocin-vasopressin system is compatible with all three of these possibilities (Smith et al., 2014).

Evidence that Empathy[1] is learned

The Learned Matching hypothesis is supported by studies of empathy in animals, infants, adults and robots.

Animals

Learned Matching predicts the occurrence of emotional contagion (Empathy[1]), not only in species where parental care and cooperation are especially important, but in all species that 1) express emotion, 2) detect the emotional expressions of others, 3) encounter distal emotional cues in a predictive relationship with experience of the same emotion, and 4) are capable of associative learning. The last of these conditions is not restrictive because associative learning has been documented in every major group of animals, including invertebrates (Heyes, 2012). The third condition also implies that emotional contagion will be found in a broad range of species because most animals experience synchronous emotion. Whenever an agent encounters a threat or exciting opportunity in the presence of conspecifics, the agent experiences an emotion alongside the opportunity to observe others experiencing the same emotion. Thus, Learned Matching is consistent with reports of emotional contagion, not only in primates (Dezecache et al., 2017; Palagi et al., 2014; Ross et al., 2008), but in birds (Osvath and Sima, 2014; Perez et al., 2015; Schwing et al., 2017; Shah et al., 2015), elephants (King et al., 2010), dogs (Huber et al., 2017; Palagi et al., 2015; Quervel-Chaumette et al., 2016, pigs (Reimert et al., 2013), rodents (Burkett, 2016; Inagaki and Ushida, 2017; Meyza et al., 2017), and ants (Hollis and Nowbahari, 2013).

Also consistent with Learned Matching, many examples of emotional contagion in nonhuman animals depend on auditory rather than visual cues, on vocal rather than facial or postural expressions of emotion (Briefer, in press). This is significant because matching emotional associations involving auditory cues are easier to learn; they can be forged by self-stimulation as well

as synchronous emotion and affect mirroring. Furthermore, Learned Matching provides a straightforward explanation for cross-species emotional contagion. It suggests that dogs 'catch' human emotions (Huber et al., 2017), and vice versa (Franklin et al., 2013), because most contemporary Western humans and their dog companions have experienced synchronous emotion. In human-canine pairs we have been startled by the same loud noises, and gladdened to see the same visitors appear at the door.

Emotional contagion has been studied most intensively in rats and mice. This research indicates that fear and pain contagion in rodents depend on the same neurological and psychological mechanisms as they do in humans. For example, in rodents as in humans, emotional contagion recruits anterior cingulate cortex; supports fear learning (inanimate stimuli encountered with a fearful target become aversive); motivates helping and consolation behaviour; and shows familiarity bias – the contagious response increases with the familiarity of the target (Burkett, 2016; Keum and Shin, 2016; Meyza et al., 2017.

Familiarity bias, and similarity bias - stronger contagious responses to targets that are morphologically similar to the agent – have been interpreted as signs that emotional contagion depends on a perception-action mechanism that is a genetic adaptation for life in "close interdependent social relationships that involve either genetic relatedness or reciprocation" (de Waal and Preston, 2017, p.503]. However, familiarity bias and similarity bias are exactly what one would expect if emotional contagion depends on Learned Matching. Familiar individuals, typically cage mates, are the targets with which the agent is most likely to have experienced synchronous emotion (and, in the human case, affect mirroring), and morphologically similar targets are more likely to produce emotional sounds, odours, and distal appendage movements resembling those produced by the agent during self-stimulation.

Further support for Learned Matching comes from evidence that rodents show contagious fear responses – freezing and squeaking when they observe a conspecific receiving electric shock – only when the observer has been shocked in the past, and has therefore had the opportunity to form matching emotional associations through self-stimulation (Church, 1959; Atsak et al., 2011).

Infants

Learned Matching suggests that, in human infants, affect mirroring – imitation by caregivers of infants' emotional displays - is an important early source of the experience that builds Empathy[1]; that constructs the associations between exteroceptive and interoceptive emotional cues enabling emotional contagion. Evidence that infants receive plenty of this kind of experience – that there is wealth, rather than "poverty of the stimulus" (Chomsky, 1975) – comes from studies showing that imitation of infants by caregivers occurs with high frequency, and commonly involves emotional displays (Gergely and Watson, 1996; Ray and Heyes, 2011). Western infants spend approximately 65% of their waking hours in face-to-face contact with caregivers (Uzgiris et al., 1989); matching behaviour occurs roughly once every minute during these interactions (Pawlby, 1977); and in 79% of cases the match results from imitation of the infant by the caregiver (Pawlby, 1977). Mothers more commonly imitate their infant's categorical emotion displays than other facial movements (e.g. twitches) (Malatesta and Izard, 1984; Malatesta et al., 1989), and their imitative behaviour is not confined to positive emotions. Mothers mirror anger and sadness as well as happiness and surprise (Tronick, 1989).

Further evidence that affect mirroring plays a key role in the development of empathy comes from research showing that the children of depressed mothers experience less affect mirroring, and are less empathic, than the children of non-depressed mothers (Field et al., 2009; Noll et al., 2012).

It is often claimed that human infants show emotional contagion, or a more sophisticated form of empathy, before they have had the opportunity to establish matching emotional associations through affect mirroring, synchronous emotion, and self-stimulation (de Waal and Preston, 2017; Meltzoff, 2011). I have not been able to find compelling evidence that this is the case. The claim rests on studies of facial gesture imitation in human newborns (Meltzoff and Moore, 1977) and contagious crying (Simner, 1971). Facial gesture imitation is not a reliable phenomenon in newborns. After years of uncertainty regarding both reliability and validity (Anisfeld, 2005; Jones, 2009; Ray and Heyes, 2011), a recent study tested more than a 100 infants, at four time points (1, 3,

6 and 9 weeks of age), for imitation of 11 gestures, using the gold standard 'cross-target' procedure (Meltzoff and Moore, 1977), and found no evidence of imitation in newborns (Heyes, 2016; Oostenbroek et al., 2016). Unlike neonatal imitation, contagious crying is a reliable phenomenon. It has been known for some 90 years that newborn human babies are apt to cry when they hear the sound of crying (Buehler and Hetzer, 1928; Simner, 1971). Studies demonstrating contagious crying in newborns are commonly cited as evidence that humans genetically inherit a propensity to feel what others are feeling (de Waal and Preston, 2017; Geangu et al., 2010; Hamlin, 2013; Liddle et al., 2015; Preston and de Waal, 2002). However, close examination of these studies points to the opposite conclusion; it supports Learned Matching.

In the first three days after birth, infants cry more when they hear the sound of another newborn crying than when they hear: equally loud white noise, background noise only, computer synthesised crying, and the cries of an infant chimpanzee (Martin and Clark, 1982; Sagi and Hoffman, 1976; Simner, 1971. These findings are consistent with a nativist view suggesting that infants genetically inherit a tendency to respond to distress stimuli with distress responses. On this view, white noise, silence and synthesised crying are less effective stimuli because they are not emotional cues, signals of distress. Given that infants cry a good deal in the first hours and days after birth, these findings are also consistent with Learned Matching – the hypothesis that infants are distressed by the sound of crying because, as a result of hearing themselves cry, they have learned an association between the sound of crying and the interoceptive experience of distress. According to Learned Matching, white noise, silence and synthesised crying elicit less crying from newborns because they are unlike the stimuli involved in learning – the sound of the infant's own cries.

Two further findings favour Learned Matching over the nativist interpretation: First, newborns cry more in response to the cries of other newborns than to the cries of older infants (Simner, 1971). This is consistent with Learned Matching because the cries of older infants are acoustically different from those of newborns (Martin and Clark, 1982), and therefore less like the stimuli involved in learning. However, it is an anomaly for the nativist view because it is not clear why a genetically inherited adaptation for empathy should discriminate against older infants

(Ruffman et al., 2017). Second, there is some evidence that newborns cry more when they hear their own pre-recorded crying than when they hear another newborn crying (Simner, 1971). This is consistent with Learned Matching because the infant's own cries are the training stimuli. However, peak responding to one's own cries is the opposite of what one would expect on the nativist hypothesis - if contagious crying is due to a genetic adaptation relating to the emotional states, not of the self, but of others.

One study found *less* crying in response to the infants' own cries than to the cries of other neonates (Martin and Clark, 1982), but it is likely that the 'other' cries in this study elicited a greater response because of their relative novelty (Dondi et al., 1999; Ruffman et al., 2017). The 'own' cries had been produced, and therefore heard, 30 secs before testing, whereas the 'other' cries had never been heard before. Thus, although contagious crying is a reliable phenomenon, the evidence suggests that it is due to the relative novelty of stimulus cries, and to matching emotional associations formed through self-stimulation.

Adults

Research with adult humans suggests that Empathy[1] depends on regions of the brain with mirror properties – including fronto-parietal motor areas, the anterior insula, and perigenual anterior cingulate cortex (Christov-Moore et al., 2014; Gallese et al., 2004; Rizzolatti and Caruana, 2017]. These regions are active both when an agent is experiencing an emotion directly, in their own right, and when the agent is experiencing the same emotion indirectly, empathically, as a result of observing a target's emotion. For example, there is evidence that the anterior insular is involved in direct experience of disgust from studies showing that electrical stimulation of the anterior insula induces disgust (Krolak-Salmon et al., 2003), and that lesions to this area impair the capacity to feel disgust in response to stimuli such as body products (Calder et al., 2000). In addition, there is evidence that the anterior insular is involved in indirect, empathic experience of disgust from studies showing that it is activated by exposure to facial expressions of disgust (Krolak-Salmon et al., 2003),

and that permanent and temporary lesions to the anterior insular impair recognition of those expressions (Calder et al., 2000; Papagno et al., 2016; Wicker et al., 2003).

Learned Matching suggests that both motoric and somatic mirror areas acquire their mirror properties through associative learning. They start out as areas involved only in direct experience of emotion; become connected through self-stimulation, synchronous emotion, and affect mirroring with areas involved in the perception of emotion cues produced by other agents; and, by virtue of these connections, end up being activated, not only directly by stimuli representing threats and opportunities for the agent, but also indirectly by the emotions of other agents (Heyes and Bird, 2007). This hypothesis, which builds on previous models of the development of mirror neurons (Cook et al., 2014; Giudice et al., 2009; Heyes, 2001; Keysers and Gazzola, 2014; Keysers and Perrett, 2004; Kilner et al., 2007), has not been tested for somatic mirror areas, such as the anterior insula and perigenual anterior cingulate cortex. However, there is now a substantial body of evidence that motoric areas in the fronto-parietal cortex acquire their mirror properties through associative learning (Cook et al., 2014). For example, research using fMRI, transcranial magnetic stimulation (TMS), electroencephalography (EEG) and behavioural methods shows that the mirrorness of motoric mirror areas – their potential to be activated by execution and observation of the same action – increases with correlated experience of seeing and doing the same action (e.g. Calvo-Merino et al., 2006; Klerk et al., 2015), and decreases, sometimes to the point of making them counter-mirror areas, with experience of seeing one action while doing another (Catmur et al., 2011; Cavallo et al., 2014).

Learned Matching predicts marked cross-cultural variation, not only in the functioning of Empathy^{2} (all theoretical perspectives would anticipate cultural variation in cognitive and metacognitive appraisal processes), but in the functioning of Empathy^{1} – in the range of emotions that are contagious, and the strength of contagious responses. For example, in cultures where there is a high rate of affect mirroring, and where children are encouraged to engage in synchronous emotion, one would expect more emotional contagion. Dedicated cross-cultural studies are needed, but, consistent with this prediction, a study involving more than 100,000 people from 63 countries

found marked cross-cultural variation in Empathic Concern, the component of the Interpersonal Reactivity Index (Davis, 1983) most closely related to emotional contagion, and greater Empathic Concern in collectivist than individualist cultures (Chopik et al., 2017).

There is a tendency to think of associative mechanisms as primitive and inflexible. Consequently, research showing that empathy is subject to modulation by contextual cues (Zaki, 2014) may appear at first sight to conflict with the Learned Matching hypothesis. However, on reflection it becomes clear that contextual modulation is at least as compatible with Learned Matching as with the standard, nativist view of the origins of Empathy[1].

In adult humans, the extent and probability of empathic responses vary with contextual factors including group membership and expertise (Hein and Singer, 2008; Zaki, 2014). Members of in-groups - political, ethnic, sports-based, and arbitrarily defined in the laboratory – provoke more empathy than members of out-groups (Avenanti et al., 2010; Cikara and Van Bavel, 2014; Hein et al., 2010; Mitchell et al., 2006); health professionals show less empathy for pain than people from other professions (Sloman et al., 2005); and people who have received meditation-based affect training show enhanced empathy / compassion (Valk et al., 2017). Many studies demonstrating contextual modulation use verbal measures that are likely to index emotion recognition rather than empathy, or a controlled response generated by the combined operation of Empathy[1] and Empathy[2]. These studies using verbal measures, although important in their own right, do not bear on the Learned Matching hypothesis because any modulation they find could be due to the cognitive and metacognitive appraisal processes of Empathy[2] (Zaki, 2014; in press).

Potentially more relevant to Learned Matching, an ERP study has found depression of responses to out-group pain just 200ms after stimulus onset (Sheng and Han, 2012). This effect occurs so early that it is unlikely to be due to appraisal of the automatic response by Empathy[2] (see Figure 2). However, early effects of this kind are compatible with Learned Matching because automatic responses mediated by learned associations are subject, not only to 'output modulation' (facilitation and inhibition following appraisal of the initial response), but also to 'input modulation' (facilitation and inhibition by attentional processes) (Heyes, 2012; Zaki, 2014). Having detected in

early trials that the target is from an out-group, the agent may pay little attention to signs of their distress, resulting in weaker activation of matching emotional associations.

Another ERP study found a weaker neural response to the observation of pain in physicians than in non-physicians just 110ms after stimulus presentation (Decety et al., 2010). In combination with evidence that medical training causes a decline in empathy (McFarlane et al., 2017), this effect of expertise can be more readily explained by Learned Matching than by nativist accounts of Empathy[1]. The development of genetically inherited adaptations is supposed to be "buffered" against environmental inputs – such as exposure to the pain of others, frequently encountered during medical training – that were present when the mechanism was evolving (Cosmides & Tooby 1994; Pinker 1997). In contrast, Learned Matching predicts alteration of automatic empathic responses, not only by input and output modulation, but by relearning or counter-conditioning (Englis et al., 1982).

Indeed, striking evidence in support of Learned Matching comes from a study showing that in human adults emotional contagion can be enhanced and suppressed using conditioning procedures (Englis et al., 1982). Applying autonomic, facial-expressive, and self-report measures, Englis and colleagues found an increase in empathic responses after congruent training – in which, for example, the sight of a target in distress was paired with experience of distress - and a decrease in empathic responses after incongruent training – in which, for example, the sight of a target in distress was paired with experience of pleasure.

The Learned Matching hypothesis suggests that the matching emotional associations mediating Empathy[1] are acquired via unbiased, domain-general mechanisms of associative learning. In principle, the learning mechanisms could be biased - we could genetically inherit a tendency to learn matching associations more readily than non-matching associations (Casile et al., 2011; Ferrari et al., 2013; Giudice et al., 2009) – but in practice I have been unable to find any evidence of such bias. For example, in the study by Englis and colleagues, after just 16 trials of incongruent training autonomic and facial expressive measures indicated either indifference to the emotional states of another agent, or "envious" and "sadistic" responding (Englis et al., 1982, p. 388).

Robots

As robots become increasingly important in industry, customer relations, and health and social care – especially care for the elderly - the artificial intelligence community has prioritised the development of empathic robots (Paiva et al., 2017). Nativist theories of empathy, including PAM (de Waal and Preston, 2017; Preston and de Waal, 2002), provide little support for this enterprise because they do not elucidate mechanisms. They *locate* empathy mechanisms – in the genes, and in parts of the brain - but they do not *explain* them; they do not tell us how the mechanisms work. Consequently, turning away from nativist models, and proving the principle of Learned Matching in a dramatic way, cognitive developmental robotics has produced artificial agents in which empathic responses to human faces and voices are based on affect mirroring, synchronous emotion, and associative learning (Asada, 2015; Lim and Okuno, 2015; Watanabe et al., 2007).

In the 'intuitive parenting' model of empathy (Figure 3), a human caregiver imitates the facial expressions of a robot baby. The robot uses a camera to sense the caregiver's facial movements, and categorises her facial expression using stored information. The caregiver's emotional expression is then connected via associative learning with the current internal state of the robot (blue lines in Figure 3). After training of this kind, the robot responds to human facial expressions of hilarity (laughter), pleasure (smiling) and displeasure (frowning) with matching facial expressions and internal states (Asada, 2015; Watanabe et al., 2007).

Concluding Remarks

Research in ethology, psychology, cognitive neuroscience and artificial intelligence – involving animals, infants, adult humans and robots – suggests that the matching mechanism at the root of empathy is assembled by associative learning. This Learned Matching hypothesis can be further tested in a variety of ways (see Figure 4). It implies that empathy's matching mechanism is functionally and anatomically specialised. The matching mechanism plays a distinctive role in producing empathic responses, depends on distinctive experience for its development, and can be

localised to particular cortical circuits. However, the Learned Matching hypothesis also implies that the matching mechanism develops and operates according to domain-general principles. For example, the development of a matching emotional association (pleasure-pleasure), like the development of a non-matching emotional association (pain-pleasure), or an association between an inanimate stimulus and a non-emotional response (bell-salivation), depends on contiguity and contingency - the associated events occurring close together in time, and in a predictive relationship.

More broadly, Learned Matching implies that empathy is not constrained by genetic evolution to favour kin and in-group members (Decety and Cowell, 2014). Cultural forces may promote bias via appraisal processes (Empathy^{2}), and, left to its own devices, associative learning will favour kin and clique members to the extent that their expressions of emotion resemble those encountered when the matching mechanism was under construction (Empathy^{1}). However, crucially, if the matching mechanism is learned it can be readily re-learned and unlearned. Even adult humans – including sectarians, medics, and people who are wary of robots - can learn to empathise more or less intensively, with a wider or narrower range of agents, not only by working on their appraisal processes (Schumann et al., 2014; Teding van Berkhout and Malouff, 2016), but by exposure to novel relationships among emotional cues (Englis et al., 1982). But this plasticity comes at a price (Heyes, 2018). Whether empathy helps or hinders morality, if the matching mechanism is learned, we cannot take it for granted that empathy will spring up with each new generation, regardless of the social environment and child rearing practices to which each new generation is exposed. Exposure to callous behaviour, in personal and public life, has the power to undermine empathy among family members, friends, colleagues and citizens.

Acknowledgements

I am grateful to Geoff Bird, Michel-Pierre Coll, and Essi Viding for their empathic education, and to Verity Brown for being a badass editor.

References

Anisfeld, M. (2005) No compelling evidence to dispute Piaget's timetable of the development of representational imitation in infancy. In *Perspectives on imitation: From cognitive neuroscience to social science (Vol.2)*, Hurley, S. and Chater, N. (eds.), MIT Press, pp. 107–131

Asada, M. (2015) Towards artificial empathy. *International Journal of Social Robotics* 7, 19-33

Atsak, P. et al. (2011) Experience modulates vicarious freezing in rats: a model for empathy. *PloS One*, 6, e21855

Avenanti, A. et al. (2010) Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology* 20, 1018-1022

Batson, C. D. et al. (2005) Empathy and altruism. In *Handbook of Positive Psychology* (Snyder, C. R. and Lopez, S. J. eds), pp. 485-498), Oxford University Press

Batson, C.D. (2009) These things called empathy: eight related but distinct phenomena. In *The Social Neuroscience of Empathy* (Decety, J. and Ickes, W., eds), MIT Press

Batson, C.D. (2011) *Altruism in Humans*, Oxford University Press

Bazelgette, P. (2017) *The Empathy Instinct*, Hodder and Stoughton

Berger, S. M. and Hadley, S. W. (1975) Some effects of a model's performance on an observer's electromyographic activity. *The American Journal of Psychology* 88, 263-276

Bird, G. and Viding, E. (2014) The self-to-other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews* 47, 520-532

Bloom, P. (2017) Empathy and its discontents. *Trends in Cognitive Sciences* 21, 24-31

Briefer, E. F. (in press) Vocal contagion of emotions in non-human animals. *Proceedings of the Royal Society B*

Bühler, C. and Hetzer, H. (1928) Das erste Verständnis für Ausdruck im ersten Lebensjahr. *Zeitschrift für Psychologie* 107, 50-61

Burkett, J. P. (2016) Oxytocin dependent consolation behavior in rodents. *Science* 351, 375–378

Calder, A. J. et al. (2000) Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience* 3, 1077–1078

Calvo-Merino, B. et al. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology* 16, 1905-1910

Casile, A., Caggiano, V. and Ferrari, P. F. (2011) The mirror neuron system: a fresh view. *The Neuroscientist*, *17*, 524-538

Catmur, C. et al. (2011) Making mirrors: premotor cortex stimulation enhances mirror and counter-mirror motor facilitation. *Journal of Cognitive Neuroscience* 23, 2352–2362

Cavallo, A. et al. (2014) Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. *Social, Cognitive and Affective Neuroscience* 9, 1082-1088

Chomsky, N. (1975) *Reflections on Language*. Pantheon Books.

Chopik, W. J. et al. (2017) Differences in empathic concern and perspective taking across 63 countries. *Journal of Cross-Cultural Psychology* 48, 23-38

Christov-Moore, L. et al. (2014) Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews* 46, 604-627

Church, R. M. (1959) Emotional reactions of rats to the pain of others. *Journal of Comparative and Physiological Psychology*, 52, 132-134

Cikara, M. and Van Bavel, J. J. (2014) The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science* 9, 245–274

Coll, M. P. et al. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews*, 83, 132-139

Cook, R. et al. (2014) Mirror neurons: from origin to function. *Behavioural and Brain Sciences*, 37, 177-241

Cosmides, L. and Tooby, J. (1994) Beyond intuition and instinct blindness: Toward an evolutionary rigorous cognitive science. *Cognition*, 50:41–77.

Davis, M. H. (1983) Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126

De Vignemont, F. and Singer, T. (2006) The empathic brain: how, when and why? *Trends in Cognitive Sciences* 10, 435-441

de Waal, F. B. and Preston, S. D. (2017) Mammalian empathy: behavioural manifestations and neural basis. *Nature Reviews Neuroscience* 18, 498-509

Decety, J. and Meyer, M. (2008) From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and Psychopathology* 20, 1053-1080

Decety, J. et al. (2010) Physicians down-regulate their pain empathy response: An event-related brain potential study. *NeuroImage* 50, 1676–1682

Decety, J. and Cowell, J. M. (2014) The complex relation between morality and empathy. *Trends in Cognitive Sciences 18*, 337-339

Decety, J. et al. (2018) The development of cognitive empathy and concern in preschool children: a behavioral neuroscience investigation. *Developmental Science*, 21, e12570. doi.org/10.1111/desc.12570

Dezecache, G. et al. (2017) Skin temperature changes in wild chimpanzees upon hearing vocalizations of conspecifics. *Royal Society Open Science 4*, 160816, doi: 10.1098/rsos.160816

Dimberg, U. (1982) Facial reactions to facial expressions. *Psychophysiology* 19, 643-647

Doherty, R. W. (1997) The emotional contagion scale: A measure of individual differences. *Journal of Nonverbal Behavior* 21, 131-154

Dondi, M. et al. (1999) Can newborns discriminate between their own cry and the cry of another newborn infant? *Developmental Psychology* 35, 418-426

Englis, B. G. et al. (1982) Conditioning of counter-empathetic emotional responses. *Journal of Experimental Social Psychology* 18, 375-391

Fan, Y. & Han, S. (2008) Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. *Neuropsychologia 46*, 160-173

Fan, Y. et al. (2011) Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews* 35, 903-911

Ferrari, P. F., Tramacere, A., Simpson, E. A. and Iriki, A. (2013) Mirror neurons through the lens of epigenetics. *Trends in Cognitive Sciences*, *17*, 450-457

Field, T. et al. (2009) Depressed mothers' infants are less responsive to faces and voices. *Infant Behavior and Development* 32, 239-244

Flom, M. and Saudino, K. J. (2017) Callous–unemotional behaviors in early childhood: Genetic and environmental contributions to stability and change. *Development and Psychopathology*, 29, 1227-1234

Franklin, R. G. et al. (2013) Neural responses to perceiving suffering in humans and animals. *Social Neuroscience* 8, 217-227

Gallese, V., Keysers, C. and Rizzolatti, G. (2004) A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*, 396-403

Geangu, E. et al. (2010) Contagious crying beyond the first days of life. *Infant Behavior and Development* 33, 279-288

Gergely, G. and Watson, J. S. (1996) The social biofeedback model of parental affect-mirroring. *The International Journal of Psychoanalysis* 77, 1181-1212

Giudice, M. D., Manera, V. and Keysers, C. (2009) Programmed to learn? The ontogeny of mirror neurons. *Developmental Science*, *12*, 350-363

Gonzalez-Liencres, C. et al. (2013) Towards a neuroscience of empathy: ontogeny, phylogeny, brain mechanisms, context and psychopathology. *Neuroscience & Biobehavioral Reviews* 37, 1537-1548

Hamlin, J. K. (2013) Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science* 22, 186-193

Hatfield, E. et al. (2014). New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion. *Interpersona 8*, 159-179.

Hawk, S. T. et al. (2012) Face the noise: embodied responses to nonverbal vocalizations of discrete emotions. *Journal of Personality and Social Psychology* 102, 796-814

Hein, G. and Singer, T. (2008) I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology*, *18*, 153-158

Hein, G. et al. (2010) Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* 68, 149–160

Heyes, C. M. and Ray, E. D. (2000) What is the significance of imitation in animals? *Advances in the Study of Behavior* 29, 215-245

Heyes, C. (2001) Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*, 253-261

Heyes, C. M. and Bird, G. (2007) Mirroring, association and the correspondence problem. In *Sensorimotor Foundations of Higher Cognition*, *Attention & Performance XX* (Haggard, P. et al., eds), Oxford University Press

Heyes, C. M. (2011) Automatic imitation. *Psychological Bulletin* 137, 463–483

Heyes, C. M. (2012) Simple minds: A qualified defence of associative learning. *Philosophical Transactions of the Royal Society B*, 367, 2695-2703

Heyes, C. (2016) Imitation: Not in our genes. *Current Biology* 26, R412-R414.

Heyes, C. M. (2018) *Cognitive Gadgets: The Cultural Evolution of Thinking*, Harvard University Press

Hoffman, M.L. (2001) *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge University Press

Hollis, K. L. and Nowbahari, E. (2013) Toward a behavioral ecology of rescue behavior.

Huber, A. et al. (2017) Investigating emotional contagion in dogs to emotional sounds of humans and conspecifics. *Animal Cognition* 20, 703-715

Ickes, W. (1997). *Empathic Accuracy*, Guilford Press

Inagaki, H. and Ushida, T. (2017) Changes in acoustic startle reflex in rats induced by playback of 22-kHz calls. *Physiology and Behavior* 169, 189–194

Jones, S.S. (2009) The development of imitation in infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 2325–2335

Kelly, J. R. et al. (2016). Emotional contagion of anger is automatic: An evolutionary explanation. *British Journal of Social Psychology* 55, 182-191

Keum, S. and Shin, H. S. (2016) Rodent models for studying empathy. *Neurobiology of Learning and Memory* 135, 22-26

Keysers, C. and Perrett, D. I. (2004) Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Sciences*, *8*, 501-507

Keysers, C. and Gazzola, V. (2014) Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philosophical Transactions of the Royal Society B*, *369*, 20130175

Kilner, J. M., Friston, K. J. and Frith, C. D. (2007) Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, *8*, 159-166

King, L. E. et al. (2010) Bee threat elicits alarm call in African elephants. *PLoS ONE* 5, e10346.

Klerk, C. C. et al. (2015) Baby steps: investigating the development of perceptual–motor couplings in infancy. *Developmental Science* 18, 270-280

Krolak-Salmon, P. et al. (2003) An attention modulated response to disgust in human ventral anterior insula. *Annals of Neurology* 53, 446–453

Lahav, A., Saltzman, E. and Schlaug, G. (2007) Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, *27*, 308-314

Lamm, C. et al. (2011) Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage* 54, 2492-2502

Liddle, M. J. E. et al. (2015) Baby empathy: infant distress and peer prosocial responses. *Infant Mental Health Journal* 36, 446-458

Lim, A. and Okuno, H. G. (2015) A recipe for empathy. *International Journal of Social Robotics* 7, 35-49

Malatesta, C.Z. and Izard, C.E. (1984) *Emotion in Adult Development*. Sage Publications

Malatesta, C.Z. et al. (1989) The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development*, i-136

Martin, G. B. and Clark, R. D. (1982) Distress crying in neonates: Species and peer specificity. *Developmental Psychology* 18, 3-9

McFarland, D. C. et al. (2017) Acute empathy decline among resident physician trainees on a hematology–oncology ward: an exploratory analysis of house staff empathy, distress, and patient death exposure. *Psycho-oncology* 26, 698-703

Meltzoff, A. N. and Moore, M. K. (1977) Imitation of facial and manual gestures by human neonates. *Science* 198, 75-78.

Meltzoff, A. N. (2011) Social cognition and the origins of imitation, empathy, and theory of mind. *The Wiley-Blackwell Handbook of Childhood Cognitive Development 2*, 49-75

Meyza, K. Z. et al. (2017) The roots of empathy: through the lens of rodent models. *Neuroscience & Biobehavioral Reviews* 76, 216-234

Mitchell, J. P. et al. (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50, 655-663

Moody, E. J. et al. (2017). Emotional mimicry beyond the face? Rapid face and body responses to facial expressions. *Social Psychological and Personality Science*. doi.org/10.1177/1948550617726832

Noll, L. K. et al. (2012) Investigating the impact of parental status and depression symptoms on the early perceptual coding of infant faces: an event-related potential study. *Social Neuroscience* 7, 525-536

Oostenbroek, J. et al. (2016) Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology* 26, 1334-1338

Osvath, M. and Sima, M. (2014) Sub-adult ravens synchronize their play: a case of emotional contagion. *Animal Behaviour and Cognition* 1, 197-205

Paiva, A. et al. (2017) Empathy in virtual agents and robots: a survey. *ACM Transactions on Interactive Intelligent Systems* 7, 11-40

Palagi, E. et al. (2014) Exploring the evolutionary foundations of empathy: consolation in monkeys. *Evolution and Human Behavior* 35, 341-349

Palagi, E. et al. (2015) Rapid mimicry and emotional contagion in domestic dogs. *Royal Society Open Science 2*, 150505, doi: 10.1098/rsos.150505

Papagno, C. et al. (2016) Specific disgust processing in the left insula: new evidence from direct electrical stimulation. *Neuropsychologia* 84, 29–35

Papousek, H. and Papousek, M. (1987) Intuitive parenting: a dialectic counterpart to the infant's integrative competence. *Handbook of Infant Development* 66, 9-720.

Parsons, C. E. et al. (2017) Intuitive parenting: understanding the neural mechanisms of parents' adaptive responses to infants. *Current Opinion in Psychology* 15, 40-44

Pavey, L. et al. (2012) I help because I want to, not because you tell me to: empathy increases autonomously motivated helping. *Personality and Social Psychology Bulletin* 38, 681-9

Pawlby, S.J. (1977) Imitative interaction. In *Studies in Mother–Infant Interaction*, Schaffer, H. (ed.), Academic Press, pp. 203-224.

Perez, E. C. et al. (2015) Physiological resonance between mates through calls as possible evidence of empathic processes in songbirds. *Hormones and Behavior* 75, 130–141

Pinker, S. (1997) *How the mind works*. Penguin Press.

Preston, S. D. and De Waal, F. B. (2002) Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*, 1-20

Quattrocki, E. and Friston, K. (2014) Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews* 47, 410-430

Quervel-Chaumette, M. et al. (2016) Investigating empathy-like responding to conspecifics' distress in pet dogs. *PLOS ONE* 11, e0152920

Ray, E. D. and Heyes, C. M. (2011) Imitation in infancy: The wealth of the stimulus. *Developmental Science* 14, 92-105

Reimert, I. et al. (2013) Indicators of positive and negative emotions and emotional contagion in pigs. *Physiology and Behavior* 17, 45–50

Rizzolatti, G. and Caruana, F. (2017) Some considerations on de Waal and Preston review. *Nature Reviews Neuroscience* 18, 769

Ross, M. D. et al. (2008) Rapid facial mimicry in orangutan play. *Biology Letters* 4, 27-30

Ruffman, T. et al. (2017) Do infants really experience emotional contagion? *Child Development Perspectives* 11, 270-274

Sagi, A. and Hoffman, M. L. (1976) Empathic distress in the newborn. *Developmental Psychology* 12, 175-176

Schumann, K. et al. (2014) Addressing the empathy deficit: Beliefs about the malleability of empathy predict effortful responses when empathy is challenging. *Journal of Personality and Social Psychology* 107, 475-493

Schwing, R. et al. (2017) Positive emotional contagion in a New Zealand parrot. *Current Biology* 27, R213-R214

Shah, S. S. et al. (2015) Risk-based alarm calling in a nonpasserine bird. *Animal Behaviour* 106, 129–136

Sheng, F. and Han, S. (2012) Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. *NeuroImage* 61, 786–797

Simner, M. L. (1971) Newborn's response to the cry of another infant. *Developmental Psychology* 5, 136-150.

Sloman, R. et al. (2005) Nurses' assessment of pain in surgical patients. *Journal of Advanced Nursing* 52, 125–132

Smith, K. E. et al. (2014) Oxytocin receptor gene variation predicts empathic concern and autonomic arousal while perceiving harm to others. *Social Neuroscience* 9, 1-9

Teding van Berkhout, E. and Malouff, J. M. (2016) The efficacy of empathy training: A meta-analysis of randomized controlled trials. *Journal of Counselling Psychology* 63, 32-41

Tronick, E.Z. (1989) Emotions and emotional communication in infants. *American Psychololologist* 44, 112-119

Uzefovsky, F. et al. (2015) Oxytocin receptor and vasopressin receptor 1a genes are respectively associated with emotional and cognitive empathy. *Hormones and Behavior* 67, 60-65

Uzgiris, I.C. et al. (1989) Contextual influences on imitative interactions between mothers and infants. In *Action in Social Context: Perspectives on Early Development*, Lockman, J. and Hazen, N. (eds.), Plenum Press, pp. 103–127

Valk, S. L. (2017) Structural plasticity of the social brain: Differential change after socio-affective and cognitive mental training. *Science Advances 3*, DOI: 10.1126/sciadv.1700489

Warrier, V. et al. (2018) Genome-wide analyses of self-reported empathy: correlations with autism, schizophrenia, and anorexia nervosa. *Translational Psychiatry*, 8, 35

Watanabe, A. et al. (2007) Mapping facial expression to internal states based on intuitive parenting. *Journal of Robotics and Mechatronics* 19, 315-320

Wicker, B. et al. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron*, 40, 655-664

Wondra, J. D. and Ellsworth, P. C. (2015) An appraisal theory of empathy and other vicarious emotional experiences. *Psychological Review* 122, 411-428.

Zaki, J. and Ochsner, K. N. (2012) The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience* 15, 675-680

Zaki, J. (2014). Empathy: a motivated account. *Psychological Bulletin* 140, 1608-1647

Zaki, J. Empathy is a moral force. In The Atlas of Moral Psychology (Gray, K. and Graham, J., eds), Guilford Press (in press)

Figure legends

Figure 1. Glossary of key terms used in this article.

Figure 2. Dual system model of empathy.

Figure 3. Robot empathy from associative learning. (Reproduced with permission from Watanabe et al., 2007).

Figure 4. Questions for future research.

Glossary

Affect mirroring: imitation by others of an agent's emotional displays; typically imitation of infants by caregivers.

Automatic processing: information processing that is fast, parallel, typically unconscious, and makes minimal demands on working memory. Within the dual system model of empathy (Box 1), automatic processing is characteristic of Empathy[1].

Canalized: the development of a characteristic, such as a psychological process, is canalized when it is invariant across a broad range of environments.

Controlled processing: information processing that is relatively slow, serial, sometimes conscious, and makes demands on working memory. Behaviour resulting from controlled processing is typically described as 'voluntary', 'intentional', or 'goal-directed'. Within the dual system model of empathy (Box 1), controlled processing is characteristic of Empathy[2].

Emotional contagion: rapid, unintentional transmission of an emotional response from one individual to another.

Empathy: I call an agent's response to a target's emotion 'empathic' if the response is caused by and resembles the target's emotion. Some researchers are more restrictive, reserving 'empathic' for voluntary responses (mediated, at least in part, by Empathy[2]), or for voluntary responses that occur when the agent believes their emotional state to have been caused by that of the agent (mediated by metacognitive as well as cognitive processes in Empathy[2]).

Learned Matching: a hypothesis suggesting that Empathy[1] depends on a matching mechanism constructed in the course of development by associative learning. The input for associative learning comes from self-stimulation, affect mirroring, and synchronous emotion.

Matching emotional association: a bidirectional excitatory link between an exteroceptive cue (e.g. a facial expression) and an interoceptive cue (motoric or somatic) belonging to the same emotional category.

Matching mechanism: a neurocognitive mechanism producing fast, emotionally congruent responses to the emotional states of other agents. Within the dual system model, the mechanism is called Empathy[1], and consists of matching emotional associations.

Metacognitive: in the dual system model (Box 1), the term 'metacognitive' refers to cognitive processes that represent the emotion-relevant cognitive processes of the self and others.

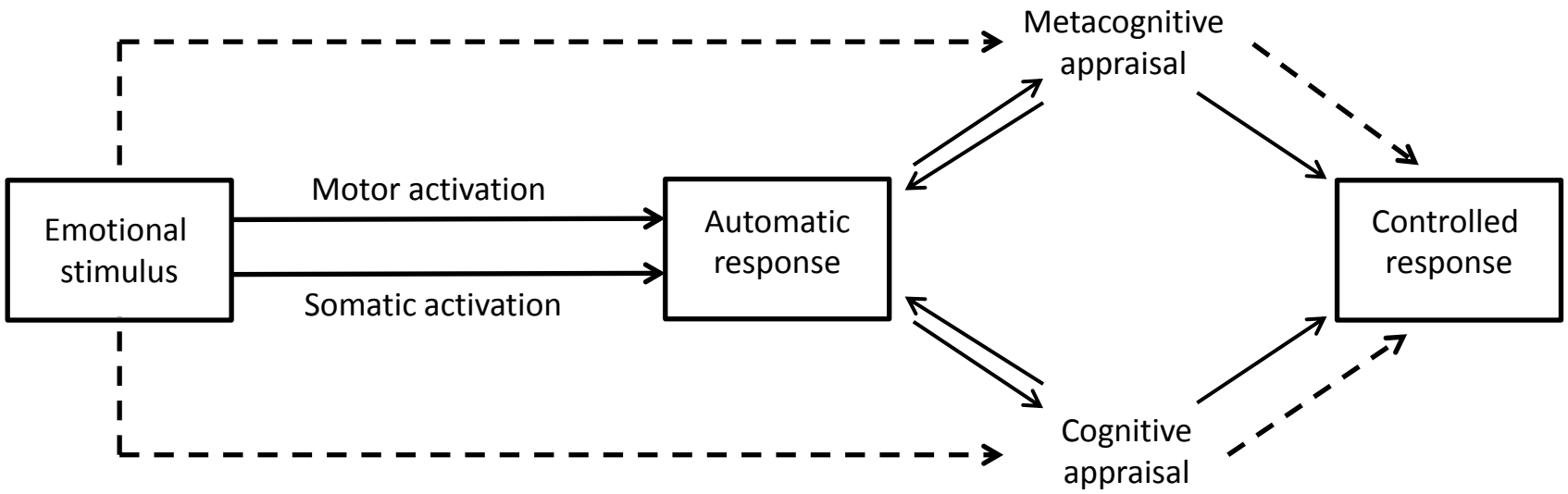
Mirror neurons: a label originally used for single neurons that discharge when the agent performs an action and when s/he passively observes the same action performed by another individual. 'Mirror neuron' and 'mirror mechanism' are now used more liberally to refer to any area of the brain that responds similarly to direct and vicarious experience.

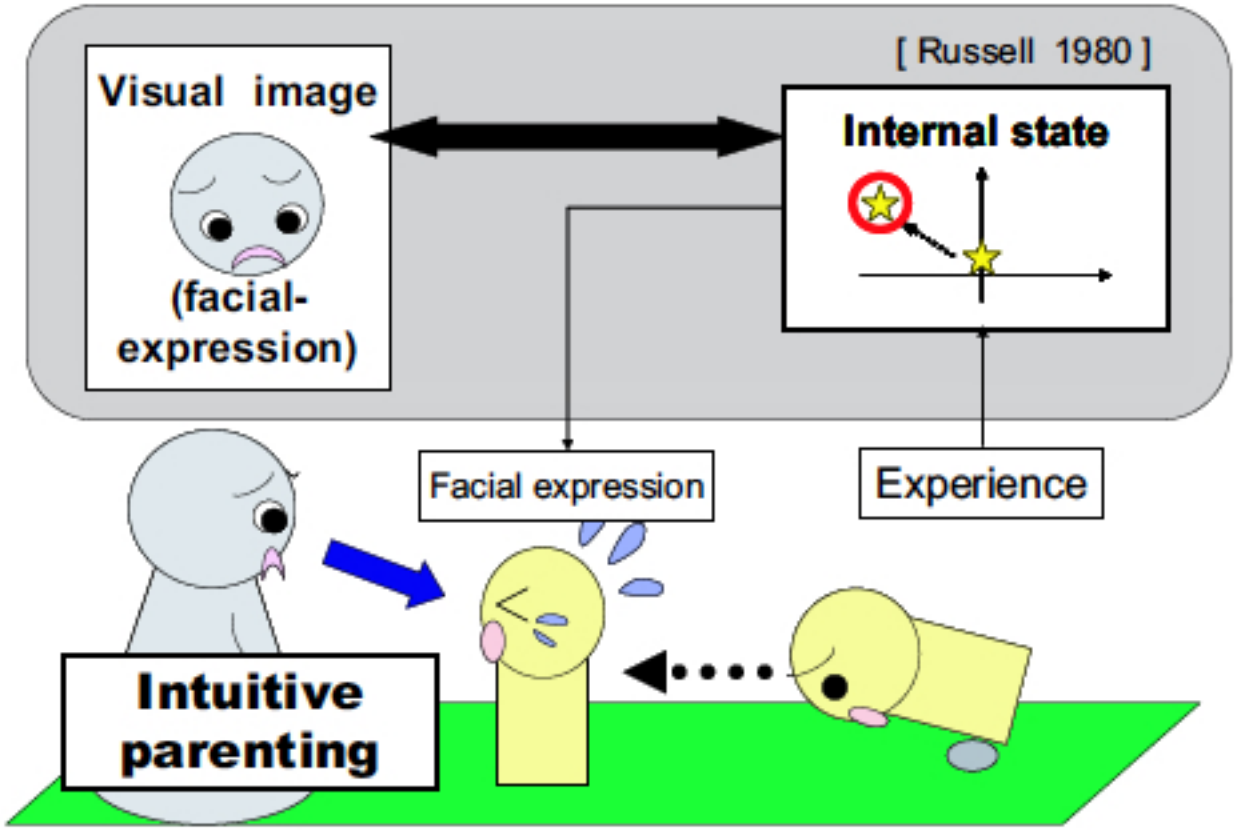
Mindreading: representing the mental states, thoughts and feelings, of the self and others. Also known as 'mentalising' and 'theory of mind'.

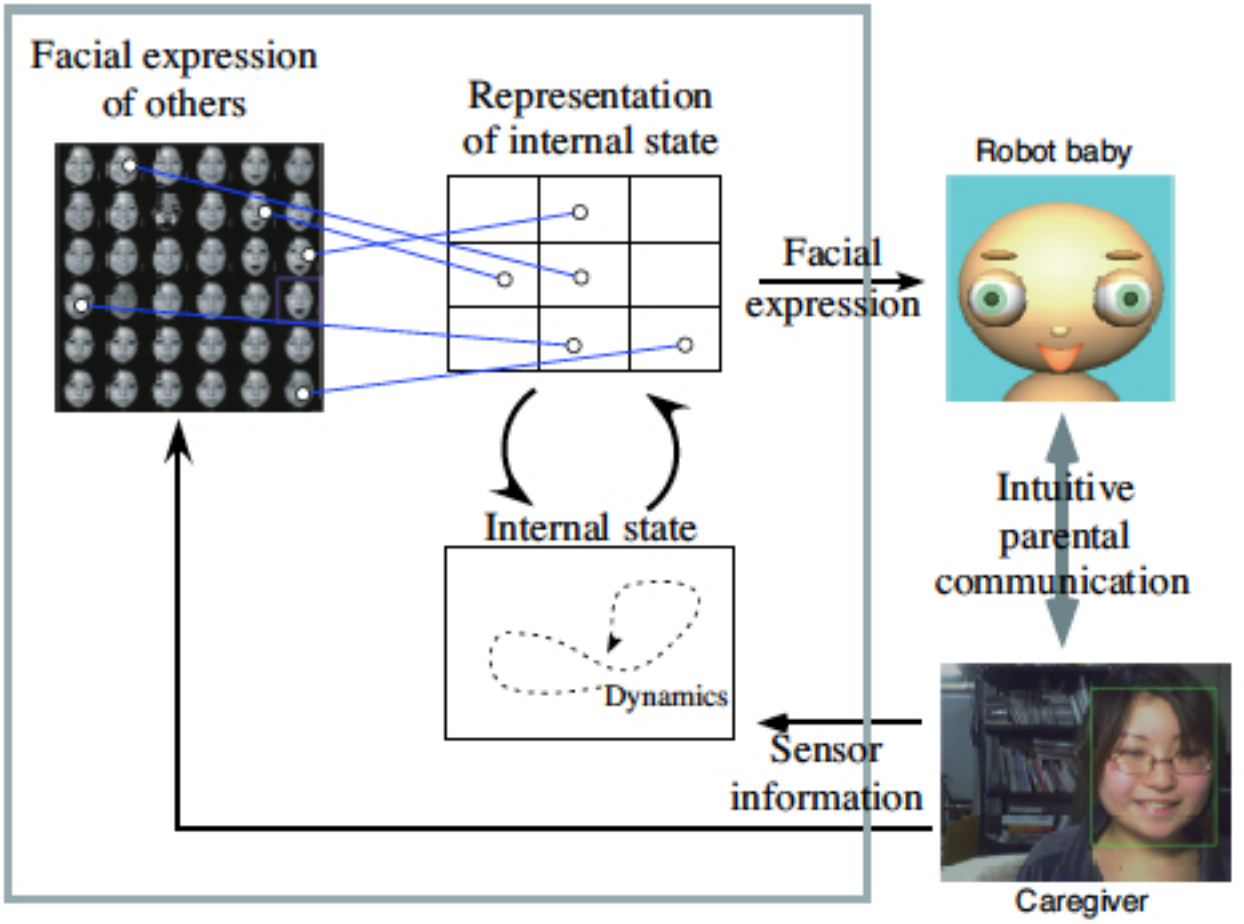
Synchronous emotion: two or more agents with perceptual access to one another exhibit similar emotional responses at the same time.

EMPATHY[1]
Emotional contagion

EMPATHY[2]
Empathic understanding




[Russell 1980]
Visual image
(facial-expression)
Internal state
Facial expression
Experience
Intuitive parenting


Facial expression of others
Representation of internal state
Robot baby
Facial expression
Internal state
Dynamics
Intuitive parental communication
Sensor information
Caregiver

Outstanding Questions

- Does emotional contagion occur in animals such as cephalopods, fishes, frogs and reptiles, where there is little cooperation and on-demand parental care? If these animals are presented with congruent emotional cues from conspecifics while they are experiencing emotion, do they begin to show emotional contagion?

- If animals are presented with incongruent emotional cues from conspecifics, do they develop counter-empathic tendencies? Are empathic and counter-empathic responses learned at the same rate?

- In infancy and childhood, are there individual differences in the degree to which facial expressions and whole body movements elicit emotional contagion? Are these individual differences selectively predicted by the frequency and sensitivity of parental affect mirroring and/or experience of synchronous emotion?

- Can typically and atypically developing children who are low on empathy be helped by interventions that provide affect mirroring and opportunities for synchronous emotion?

- To what extent do cultures vary in the range of emotions that are contagious and the strength of contagious responses? Do these cultural differences co-vary with practices that encourage affect mirroring and provide opportunities for synchronous emotion?

- Does between-group experience of affect mirroring and synchronous emotion counteract in-group bias in empathic responding?

- Do people show stronger automatic empathic responses to artificial agents when they have been affectively mirrored by those agents?