Are we Really Measuring Empathy? Proposal for a New Measurement Framework.

Michel-Pierre Coll[1,2,*], Essi Viding[3], Markus Rütgen[4], Giorgia Silani[5], Claus Lamm[4], Caroline Catmur[6] and Geoffrey Bird[1,2]

1 Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.

2 MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, University of London, SE5 8AF, UK.

3 Developmental Risk and Resilience Unit, Division of Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AP, UK.

4 Social, Cognitive and Affective Neuroscience Unit, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Vienna 1010, Austria.

5 Department of Applied Psychology: Health, Development, Enhancement and Intervention, University of Vienna, University of Vienna, Vienna 1010, Austria.

6 Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College, London, London, SE5 8AF, UK.

*Correspondence to: michel-pierre.coll@psy.ox.ac.uk

Abstract

Empathy - currently defined as the sharing of another's affective state - has been the focus of much psychological and neuroscientific research in the last decade, much of which has been focused on ascertaining the empathic ability of individuals with various clinical conditions. However, most of this work tends to overlook the fact that empathy is the result of a complex process requiring a number of intermediate processing steps. It is therefore the case that describing an individual or group as 'lacking empathy' lacks specificity. We argue for an alternative measurement framework, in which we explain variance in empathic response in terms of individual differences in the ability to identify another's emotional state ('emotion identification'), and the degree to which identification of another's state causes a corresponding state in the self ('affect sharing'). We describe how existing empathy paradigms need to be modified in order to fit within this measurement framework, and illustrate the utility of this approach with reference to examples from both cognitive neuroscience and clinical psychology.

Keywords: Empathy; affect sharing; emotion identification; neuroscience; model; theory; definition.

Highlights

- Empathy is currently defined as sharing the state of another.
- This definition conflates the identification and the sharing of another's state.
- Describing change or impairment in empathy therefore lacks specificity.
- We show how this can be problematic for popular paradigms in social neuroscience.
- And propose an alternative measurement framework to resolve this issue

1 Introduction

Empathy is commonly understood to be a complex psychological construct that plays a crucial role in social interaction. As with many complex constructs, several overlapping but distinct definitions of empathy have been suggested (Batson, 2009; Cuff, Brown, Taylor, & Howat, 2016). While there is as yet no consensus as to the precise definition of empathy, most researchers (at least in the field of cognitive neuroscience and psychology) agree that empathy involves the adoption of another's affective state so that both the Empathizer and the empathic target (henceforth 'Target') are in a similar state (Cuff et al., 2016; Decety & Jackson, 2004; de Vignemont & Singer, 2006; de Waal, 2008; Zaki & Ochsner, 2012; Shamay-Tsoory, Aharon-Peretz & Perry, 2009). This notion of sharing the affective state of another forms the core of what we shall refer to as the standard definition of empathy.

Empathy has received considerable research attention in the last decade, with a particular focus on its neural instantiation permitted by improvements in human functional neuroimaging (Lamm, Bukowski, & Silani, 2016; Shamay-Tsoory, 2011; Singer & Lamm, 2009; Zaki & Ochsner, 2012). Establishing the neural networks underlying empathy can elucidate the relationship between self- and other-related affective experiences, provide information about the functional processes involved in empathy, and suggest interventions to modulate levels of empathy wherever desired.

Despite several leading theoretical models arguing for a multi-factorial structure of empathy (Davis, 1980; Decety & Jackson, 2004; Decety & Meyer, 2008; Preston & de Waal, 2002), there have been surprisingly few efforts to develop exhaustive information processing models to detail the different processing stages involved in producing an empathic response. One

consequence of this is that it becomes difficult to determine the locus of any effect that influences the empathic response. Without consideration of the contribution of those processes upon which empathy relies, one cannot be sure that any effect is on empathy *per se*, or on a computational precursor. Here, it will be argued that empathy relies upon, but is distinct from, the ability to identify the emotional state of the Target (Bird & Viding, 2014; Happé, Cook, & Bird, 2017). The implication of this distinction between empathy and emotion identification for past and future research will be discussed by showing that failing to distinguish these two constructs could interfere with the correct interpretation and measurement of differences in empathic responses associated with experimental manipulations or clinical conditions. Distinguishing between emotion identification and empathy necessarily requires refinement of at least the standard measurement framework for empathy, and possibly the definition of empathy itself.

2 Current issues with the measurement of empathy

Under the standard definition, for empathy to have occurred, the Empathizer must be in a similar affective state to the Target. It therefore necessarily follows that in order to demonstrate an empathic response, the Empathizer must be able to identify the Target's affective state accurately, and identification of the Target's state must cause the Empathizer to share this state. Under the standard definition of empathy therefore, the Empathizer can only be considered empathic if they correctly **identify and share** the Target's emotion. Conversely, in cases where the Empathizer does not identify the Target's state accurately, irrespective of whether the Empathizer shares the state they judge the Target to be in, they cannot fulfil the standard definition of empathy (Bird & Viding, 2014).

Under the standard definition then, empathy is a state one enters into as a consequence of at least two processes (emotion identification and affect sharing): and empathy is just one possible outcome of these two processes (for example, any inaccuracy of emotion identification will result in a non-matching state). How then, should one conceptualize individual differences in empathy? If empathy refers to the outcome of two processes, and, if either of these processes is not functioning perfectly the outcome does not meet the definition of empathy, then what does it mean to be less empathic? It is true that the affective state which arises as a consequence of these two processes can be more or less like the state of the other. However, any state deviating from the matching state does not meet the definition of empathy. Under the standard 'matching state' definition therefore, empathy is binary – it either occurs or does not. This definition is incompatible with the common understanding of empathy, in which it is acknowledged that there can be varying degrees of empathy and that individuals or groups can be more or less empathic. Despite this, we shall continue to use the term empathic response to refer to the outcome of the emotion identification and affect sharing processes as it is the term most commonly used in the literature.

As can be seen then, to describe an individual or group as 'less empathic' is problematic when empathy is defined as a state. However, even if this problem is overlooked, the fact that empathy is the product of two processes means that one can be 'less empathic' either because one has misidentified the Target's state, or because even though the Target's state has been correctly identified, one does not share the Target's state. This is an unsatisfactory situation as, according to current usage, the notion of 'impaired empathy' conflates two processes: the identification of the Target's state, and the sharing of the Target's state. These processes

contribute independent variance to the empathic response and can be independently affected in clinical conditions. Furthermore, it is likely that a clinical group characterized by reduced empathy due to poor emotion identification will need a different intervention than a group also characterized by reduced empathy, but where this is due to reduced affect sharing.

It seems that there are two possible solutions to this problem: The first is that we continue to use the standard definition of empathy as the outcome of two processes but we do not refer to individual or group differences in empathy; rather we specify whether any individual differences, experimental manipulations, or clinical conditions impact emotion identification, affect sharing or both. This solution has the benefit of keeping the standard definition of empathy, but dissociates the concept of empathy from measurement of the processes giving rise to the empathic response. The second solution is to redefine empathy such that rather than the outcome of a process it becomes the process of affect sharing itself; however, it would be measured not as the degree to which the Empathizer's state matches that of the Target, but rather the degree to which the Empathizer's state matches that identified in the Target (which may deviate from the Target's actual state). This solution has the benefit that it becomes meaningful to discuss individual differences in empathy (because empathy is no longer binary), and individual differences in empathy are directly related to the measurement of a single process rather than a conflation of two processes. A drawback of the new definition is that it deviates both from the long tradition of existing work on empathy using the standard definition, and from the popular understanding of empathy. While either approach is logically coherent, it should be noted that the implications for the measurement of empathy that are outlined below are the same whichever option is chosen. The first solution is relatively easy to implement and the

118 section *"Implications for paradigms used in basic and clinical studies of empathy"* will describe
119 how this can be done within existing empathy paradigms. The second solution is more radical
120 and therefore we have not pursued it further here, but note that adoption of this definition may
121 be worthy of consideration by the field in future.
122

123 3 Defining and measuring emotion identification and affect 124 sharing

125 We consider emotion identification to be the process of attributing an emotion to an
126 individual (note that this need not be a conscious attribution) which is agnostic as to the method
127 by which the attribution is made - it can be based on observable perceptual cues, but also
128 includes identification of an individual's state based on contextual information or inferential
129 reasoning. As such, it encompasses the stages of emotion perception, recognition and
130 categorization (see Schirmer & Adolphs, 2017 for a definition of these concepts). The accuracy of
131 emotion identification is therefore defined as the degree to which the Empathizer's judgement
132 of the state of the Target matches the Target's actual state (Figure 1- top panel). Specific methods
133 for measuring emotion identification are outlined below, but it is immediately apparent that an
134 individual may vary in their ability to identify another's emotion depending on the cues available
135 to them and on the context the Target is in. For example, an Empathizer with a specific problem
136 with the recognition of emotional facial expressions may be very inaccurate in identifying the
137 Target's state when the Target's facial expression is the only information the Empathizer has to
138 make their judgement, but be much more accurate if they know the situation the Target is in and

have been in a similar situation. The processes contributing to emotion identification will also be recruited to explain and predict behaviour without necessarily evoking an emotional response, but here we are interested in their role in producing an empathic response.

Affect sharing describes the process whereby identification of another's state causes that state to be instantiated in the self. Individual differences in affect sharing would be described by differences in the function mapping the state elicited in the Empathizer as a result of their judgement of the Target's emotional state (not the Target's actual state; see Figures 1 [bottom panel] and 2). For example, if the affect sharing function can be described as a simple ratio (note that more complicated functions are possible, and even probable - see Figure 2), then an individual with a ratio of 2:1 (emotion identified in the other : emotion elicited in the self), would be described as having a greater degree of affect sharing than an individual for whom the ratio is 3:1. This is because, given that they both identify the same state in the Target, the state elicited in the former individual will be greater than the state elicited in the latter individual. Affect sharing may be described as more or less accurate on the basis of the degree of correspondence between the state of the Target identified by the Empathizer and the empathic response elicited in the Empathizer. A high degree of correspondence indicates a high degree of accuracy, whereas 'too much' affect sharing is indicated when the state elicited in the Empathizer by their judgement of the Target's state is more extreme than the state attributed to the Target (described by a ratio of 1:2 using the example above). This ratio describes affect sharing ability independently of potential differences in emotion identification, such that individuals with a similar ratio can be deemed to have similar degrees of affect sharing regardless of their ability to identify another's emotion.

Using these conceptualizations of emotion identification and affect sharing, an empathic response (state) is the product of emotion identification and an individual's degree of affect sharing; separate processes that contribute independent variance. For two individuals who have the same degree of affect sharing, i.e. their empathic response will be identical given that they identify the same affective state in another, any difference in their empathic response will reflect differences in their judgement of the Target's emotion (emotion identification). Conversely, for two individuals equally good at identifying the state of the Target, any difference in the degree of empathic response elicited will be due to differences in their degree of affect sharing (Figure 3).

The importance of measuring, and distinguishing between, emotion identification and affect sharing, is illustrated by the following, somewhat artificial, thought experiment. Consider the case of a parent who sees their child injured and in great pain, and consequently feels a great deal of empathic pain on their behalf. If the same parent on a different occasion sees the child suffer a minor misfortune resulting in only temporary and mild pain, and feels an empathic pain response that is reduced compared to that which they felt on the first occasion, then one would not infer that the parent had become less empathic (or more formally that their degree of affect sharing had reduced). One would infer that their degree of affect sharing remained the same and that their empathic response was appropriate for the degree of pain attributed to their child in the latter case, even though their empathic response was reduced. On a within-subject level therefore, one cannot assume that a reduced empathic response observed at a certain time point, or after a specific manipulation, is an indicator that affect sharing itself has been reduced unless it can be demonstrated that the empathic response is less than expected given the state

identified by the Empathizer in the Target.

The same logic holds for between-subjects comparisons. To return to our thought experiment, let us consider the case of two adults who see a child undergo an innocuous accident which would cause only mild and temporary distress in the vast majority of children. However, one of the adults knows that the child suffers from juvenile arthritis and will therefore experience a large degree of pain. We would not infer that the greater degree of empathic pain experienced by this adult is a result of them being more empathic (more formally that they had a greater degree of affect sharing) than the adult who is ignorant of the child's condition. Rather, we would explain their greater empathic response with the fact that they have identified a higher degree of pain in the child.

These thought experiments illustrate that in order to correctly measure an individual's degree of affect sharing, one can neither rely solely on the Empathizer's empathic response, nor on the accuracy with which they can identify the Target's affective state, but must instead use the degree of correspondence between the Empathizer's empathic response and the Empathizer's identification of the Target's state. Without measurement of both of these factors, it is impossible to dissociate emotion identification and affect sharing in order to explain variance in the empathic response.

4 Implications for paradigms used in basic and clinical studies of empathy

We have argued that it is necessary to distinguish between emotion identification and affect sharing in order to characterise individual differences in the empathic response, unless empathy is redefined as affect sharing. However, whether one retains the existing definition of empathy but measures differences in emotion identification and affect sharing, or adopts the new definition of empathy, the methodological implications are identical: one must obtain independent measures of emotion identification and affect sharing. This new methodological framework has important implications for the most commonly used measures of empathy, and for the interpretation of manipulations aimed at modulating empathy. Several of these are outlined below, with discussion of how methods or interpretations may need revising in light of the distinction between emotion identification and affect sharing.

4.1 The Empathic Accuracy Task (EAT)

The EAT, based on work by William Ickes and others (Ickes, Stinson, Bissonnette, & Garcia, 1990; Levenson & Ruef, 1992) and subsequently used by Zaki and colleagues (Devlin, Zaki, Ong, & Gruber, 2016; Zaki, Bolger, & Ochsner, 2008) is a measure in which a group of interviewees (Targets) describe an emotional experience while providing continuous ratings of how they feel. These videotaped interviews are then used as stimulus material for experimental participants (Empathizers); while watching the videos the participants are asked to provide continuous ratings of the emotional state of the interviewee. Traditionally, the data are analysed

by calculating the degree of correlation between the continuous ratings provided by the interviewee and those provided by experimental participants. This degree of congruence is described as a measure of empathic accuracy. However, based on the framework described above, we would suggest that any discrepancy between the ratings provided by the interviewee and the experimental participant might be better characterized as an error in emotion identification. We can see from Figure 1 that Individual D would be described as having perfect empathic accuracy on this measure, even though they lack any empathic response to the state of the other. Accordingly, a valuable addition to this task, and, as will become apparent, to all empathy tasks, would be to require participants to provide two sets of ratings: the first, as used in the existing version of this task, indicating how they think the interviewee feels; and the second indicating how they themselves feel. When these two sets of ratings are obtained, the participant's judgement as to the state of the interviewee and the interviewee's report of their own state can be compared to obtain a measure of the accuracy of emotion identification, whereas the participant's judgement of the interviewee's state and the participant's report of their own state can be compared to derive a measure of affect sharing (as described in Figure 2). Ideally, steps should be taken to avoid these ratings influencing each other. For example, the ratings could be obtained during separate experimental sessions in a counterbalanced order.

4.2 'Implicit empathy' paradigms

First utilized by Jackson, Meltzoff, & Decety (2005), these paradigms involve the participant being presented with images of bodies in either painful or non-painful situations. The neural activity elicited by the painful images is compared with that elicited by the non-painful

images in order to obtain a neural signature of empathy-related brain activity. This activity can then be compared across individuals or groups. A behavioural variant of this procedure was used by Gu and collaborators (Gu et al., 2010) in which participants were asked to perform an incidental task (e.g. determining whether images were of a left or a right hand) with the same painful and non-painful images of body parts. Reaction times on the incidental task were compared for painful and non-painful images on the assumption that images of others in pain would interfere with performance on the incidental task due to the empathic distress they evoke, and that therefore the degree of interference (in terms of reaction time) is an index of empathy.

As noted above, however, with these paradigms it is not clear how much of the variance in the empathic response (whether behavioural or neural) is due to variance in emotion identification, and how much to affect sharing. In the original study by Jackson et al. (2005), there was a strong correlation between the intensity of pain identified in the other and activation in the mid cingulate cortex, an area often, but not always, associated with empathy (see Lamm, Decety, & Singer, 2011 for a meta-analysis). Although pain intensity ratings might be influenced by both emotion identification and the empathic response, it is likely the case that a substantial proportion of the variation in the empathic brain response is due to variation in emotion identification. Therefore, when these paradigms are used to compare the effect of experimental manipulations or group membership, effects on the degree of pain identified in the other should be measured and taken into account in the analysis of any effect on the empathic response - if changes in identified pain fully explain changes in the empathic response then the effect can be attributed to emotion identification, whereas if an effect persists after accounting for variance in emotion identification then one can be more confident in attributing any effect to affect sharing.

For example, Decety and collaborators (Cheng et al., 2007; Decety, Yang, & Cheng, 2010) showed that physicians had a decreased neural response when observing pain in others. As in Jackson and collaborators' study, the neural response was correlated with pain intensity ratings, which were lower in the group of physicians. Therefore, it is possible that the difference in the neural response in physicians is solely due to differences in pain identification – indeed, the underestimation of the intensity of patients' pain in physicians is well-documented (see Prkachin, Solomon, & Ross, 2007 for a review). If this is the case, then matching the stimuli presented on the degree of pain identified by each group should lead to typical empathic responses in physicians.

4.3 'Shared Network' imaging studies with a fixed stimulus

One of the first neuroimaging studies of empathy was performed by (Singer et al., 2004). This study is of particular interest as participants were only shown one of four coloured arrows. Each arrow signalled that either the participant or their experimental partner would receive a painful or a non-painful electric shock (each arrow signalled one of the four possibilities). Crucially, before the experiment, both the participant and the partner underwent a pain thresholding procedure so that all participants received a shock calibrated to produce a fixed percentage of the maximum pain they could tolerate. Thus, in principle, receipt of the painful shock had the same subjective value for the participant and their partner. Following this procedure, the degree of empathic brain activity elicited by the partner's painful shocks in areas of the brain responding when participants received pain themselves served as a neural index of empathy. Despite only measuring the empathic response, the use of this paradigm is less

susceptible to interference from variance in emotion identification as, at least in principle, the degree of pain is fixed for all participants. While a manipulation check could have been used to ensure that there wasn't variance in the extent to which participants judged their partner to have habituated to the shock, or the extent to which their partner may have experienced increasing pain summation with repeated shocks, the use of a clearly defined and unchanging Target pain intensity is of value here. The implication is that any variance in the neural empathic brain response is attributable to affect sharing rather than emotion identification.

A variant of this procedure was used in two of our (M.R., G.S. and C.L.) recent studies (Rütgen, Seidel, Riečanský, & Lamm, 2015; Rütgen, Seidel, Silani, et al., 2015). These experiments were designed to investigate the effect of reduced self-pain on the response to the pain of others. The participant's own pain was reduced with use of a placebo analgesia procedure in which the participants were given an inert pill and informed that it would reduce their pain. In common with previous demonstrations of placebo analgesia this manipulation was successful; electric shocks were perceived as less painful than prior to the manipulation. The Singer and colleagues paradigm described above was then administered, with the addition of a photograph of the partner's pained facial expression when they received a shock. Crucially, participants were asked to judge the degree of pain experienced by the partner when the partner received a shock, and also how bad the partner's shock made the participant feel. Although not the focus of either paper, the fact that participants were asked to report their estimate of their partner's pain, and the degree of affective response evoked in themselves, enable the independent effects of the intervention on emotion identification and affect sharing to be established.

The results obtained by Rütgen and colleagues show that the placebo analgesia

manipulation reduced self-reported empathic responses (i.e., “How unpleasant did it feel when the other person was stimulated?”), but also reduced the intensity of the pain perceived in the partner (i.e. “How painful was this stimulus for the other person?”). Since the decrease in the participant’s empathic response was similar to the reduction in the intensity of the pain perceived in their partner, it is therefore possible that the effect of the placebo analgesia manipulation on the empathic response is solely a product of the reduction in the intensity of the perceived pain (i.e. an effect on emotion identification), and not explained by an effect on affect sharing. This was supported by the results of a mediation analysis on the original behavioural data from the 102 participants reported in the Rütgen, Seidel, Silani, et al. (2015) study. The results obtained (Figure 4B) show that the effect of the placebo analgesia manipulation on the empathic response (ratings of how unpleasant it was for the self when the other received pain) was fully mediated by the intensity of the pain attributed to the partner (intensity of other-pain ratings). Indeed, a significant indirect effect (ab = 0.46, bootstrap 95% confidence interval: 0.11-0.79) explained 93% of the effect of the placebo manipulation on the empathic response. These data thus suggest that the effect of the placebo analgesia manipulation was on emotion identification and not affect sharing.

These data also allow an alternative model to be tested; that there is a feedback effect of the empathic response on emotion identification. Note that emotion identification would still contribute unique variance to the empathic response – the state identified in the Target would determine, in part, the empathic response elicited in the Empathizer, and therefore would constrain the degree of empathic response available to modulate emotion identification. The feedback model can be tested using the data of Rütgen, Seidel, Silani, et al., (2015) by assessing

the mediating effect of the empathic response on the relationship between the placebo manipulation and emotion identification (Figure 4C). This analysis showed that the indirect effect in the mediation model did not reach significance (indirect effect ab = 0.30; bootstrap 95% confidence interval: -0.02-0.66), and that although the empathic response explained 54% of the placebo effect on emotion identification (compared to 93% of the effect explained by the emotion identification mediation model), the placebo manipulation was still a significant predictor of emotion identification after the empathic response was taken into account (path c', p = 0.035, one-tailed). It should be noted, though, that the two types of ratings were not counterbalanced; other pain estimates were always collected before ratings of the empathic response. These results should therefore be interpreted with caution due to the possible presence of an order effect. While these findings do not therefore necessarily imply that placebo analgesia always exerts its effects on empathy by influencing emotion identification alone, they are used here to illustrate the importance of considering emotion identification and affect sharing as processes that can vary independently.

The inclusion of measures of both empathic response and emotion identification is a useful feature of the Rütgen and collaborators studies. Other studies aiming at manipulating empathy did not follow this procedure and therefore cannot distinguish between changes in emotion identification and affect sharing. For example, recent neurostimulation studies (including one from our group (M.-P.C.)) have interpreted changes in intensity ratings of others' pain following transcranial direct current stimulation to the dorsolateral prefrontal cortex (Wang, Wang, Hu, & Li, 2014) or the temporoparietal junction (Coll, Tremblay, & Jackson, 2017) as changes in empathic responses. However, in both of these cases, since empathic responses were

not measured, it could be the case that the stimulation only altered the participants' emotion identification.

This brief review of empathy paradigms and empathy modulation studies further illustrates that changes in affect sharing should be measured as changes in the relationship between the intensity of the emotion attributed to the Target and the degree of the empathic response to the Target's state (Figure 3). Alternatively, a mediation model may be used in order to determine whether emotion identification mediates the effect of any intervention on the empathic response: If emotion identification fully mediates any effect on the empathic response then it is likely that affect sharing is not affected. The important implication of this empirical framework is that we should no longer talk of modulations of empathy, rather we should distinguish between modulation of emotion identification and affect sharing (or, as mentioned above, redefine empathy as affect sharing). A claim that a manipulation affects affect sharing should be accompanied by a demonstration that any modulation of the empathic response is independent of (or at least not fully explained by) altered emotion identification. This can be achieved by measuring and taking into account emotion identification when testing empathic responses, or by individually calibrating the stimuli used to ensure that all participants attribute the same degree of emotion to the Target. Future studies should also further assess the typical relationship between emotion identification and affect sharing across the population and the factors that can influence this relationship.

5 Affect sharing and emotion identification in clinical conditions

Due to its crucial role in social interaction, there has long been an interest in assessing

empathy in clinical conditions thought to be characterized by impaired social functioning. In recent years this has led to the frequent use of the paradigms discussed above, and other approaches, to measure empathic responses in clinical populations. While it is beyond the scope of this paper to describe how emotion identification could explain many findings suggesting altered empathy in clinical populations, the distinction between emotion identification and affect sharing has important implications for future clinical research on empathy. For example, there is accumulating evidence that levels of alexithymia, a sub-clinical condition associated with problems in identifying one's own emotions (Nemiah, Freyberger, Sifneos, & Others, 1976), can explain the poor ability to identify the emotion of others which is observed in several psychiatric disorders (Bird & Cook, 2013; Brewer, Cook, Cardi, Treasure, & Bird, 2015; Cook, Brewer, Shah, & Bird, 2013; Heaton et al., 2012; Keysers & Gazzola, 2014; Lamm et al., 2016). Therefore, the investigation of empathic responses within these clinical groups should describe potential differences in empathy in relation to the ability to identify one's own emotional states and the emotional states of others. Adequately characterizing each of these abilities and their interaction will help improve future research and psychological treatments. This is especially important as it is likely that deficits in emotion identification will require different therapeutic interventions than those designed to increase affect sharing, even though both interventions may result in an increased empathic response. With respect to a condition such as psychopathy, for example, we have previously argued that psychopaths may have impaired emotion identification, and this is, in part, why they do not develop typical affect sharing. Investigating emotion identification and affect sharing longitudinally in young children with psychopathic traits would help confirm whether this proposition is correct. In contrast, those with Autism Spectrum Disorder may have

intact affect sharing and emotion identification when contextual and social inferences are not necessary (Bird & Viding, 2014; Fan et al., 2013; Hadjikani et al., 2014; Lockwood, Bird, Bridge & Viding, 2013; Tell & Davidson, 2014).

6 Further considerations

Although we have argued for the independence of emotion identification and affect sharing, it is clear that this is an oversimplification of the complete empathic process (see Bird & Viding, 2014 for a more comprehensive attempt to identify all the processes involved in generating an empathic response). There are many processes that may impact on the empathic response, and on emotion identification, that are not addressed here (these include action perception, theory of mind, and interoception) and all may make the relationship between emotion identification, affect sharing, and the empathic response difficult to observe in experimental settings if they do not include the means to experimentally or statistically account for variance in these additional processes. It should also be recognized that the empathic response is the result of a dynamic process which unfolds over time, with the possibility of recurrent processing and feedback from later processing stages to earlier processing stages. Rather than negate the necessity of dissociating emotion identification and affect sharing, considerations such as these highlight that the ultimate aim should be to produce a dynamic model of all processes that contribute to the empathic response in order to gain a complete picture of an individual's or group's socio-affective ability, or to understand the impact of an intervention which modulates the empathic response.

It should also be noted that we have not addressed the distinction which is sometimes

made in the literature between empathy and emotion contagion (e.g. de Vignemont & Singer, 2006; de Waal, 1996; see also Hatfield, Cacioppo, & Rapson, 1993). This distinction is typically drawn on the basis of self-other distinction; for example de Waal (1996) defines emotional contagion as *"total identification without discrimination between one's feelings and those of the others* (p. 80)*"* whereas empathy occurs when *"the other is recognized not just as an extension of the self, but as a separate entity* (p. 69)*".* Singer & de Vignemont (2006) go further, stating that empathy is distinguished from emotion contagion when the Empathiser realises that their state has been caused by the state of the Target. The distinction between emotion contagion and empathy is clearly important for the phenomenology of the empathic experience; and influences whether the Empathiser feels a state of personal distress due to a lack of self-other distinction between their state and the negative state of the Target, or a state of empathic concern (Nancy Eisenberg & Sulik, 2012). It also likely influences the likelihood and type of behaviour in response to another's state - personal distress may prompt a withdrawal response from the Target, whereas empathic concern is more likely to prompt prosocial helping behaviour (Batson, Fultz, & Schoenrade, 1987; de Waal, 2008; Eisenberg, Hofer, & Vaughan, 2007). However, this distinction has less relevance for the framework presented above. As previously noted, emotion identification may involve a conscious recognition of the state of the Target or not. If emotion identification is accurate but not conscious, and the affect sharing system is intact, then emotion contagion (as defined by de Waal, 1996) will result. If emotion identification is accurate and conscious, and the affect sharing system is intact, then the Empathiser will be in the same state as the Target (meeting the standard definition of empathy) and will have a conscious representation of the Target's state. It is an open question as to the factors that determine

whether the Empathizer then engages in self-other distinction (Bird & Viding, 2014; de Waal, 2008), or realizes that their state has been caused by that of the Target (de Vignemont & Singer, 2006).

Finally, the main focus of this text has been on empathy in humans and we have not addressed the implications of this new framework for non-human animal studies, which can provide an important contribution to the understanding of the cognitive and affective processes underlying empathy (de Waal & Preston, 2017; Panksepp & Panksepp, 2013). Since it has been previously argued that empathy is supported by similar processes in nonhuman mammals (Meyza, Bartal, Monfils, Panksepp, & Knapska, 2017; Panksepp & Lahvis, 2011), it would be interesting for future studies to also attempt to measure and dissociate processes akin to emotion identification and affect sharing in non-human animals.

7 Conclusion

The fact that emotion identification and affect sharing are often confounded in experimental paradigms, or used as interchangeable terms, or described as 'empathy' reflects both the paucity of information processing models of socio-cognitive processes and the lack of a common lexicon in the social cognition literature (Happé et al., 2017). These processes may be interrelated, but they need to be considered independently to understand the mechanisms underlying individual differences in empathic responses, and to identify the locus of any modulation of empathic response in clinical populations or due to psychological or pharmacological interventions. Adequately characterizing each of these mechanisms and their interaction will help improve future cognitive neuroscience research and psychological treatments. Furthermore, and equally important, consideration of the differential impact of

impaired emotion recognition and affect sharing leads us to offer a novel empirical framework to measure empathy, and to describe variance in empathic responses. Whether this aim would be better served by redefining empathy as affect sharing - as the degree to which the Empathizer's own state matches that identified in the Target - is an open question. Nevertheless, the resolution of this issue does not negate the requirement to measure emotion identification and affect sharing independently in any study of empathy, and so we recommend the use of the measurement framework described here.

Conflict of interest

The authors declare no competing financial interests.

Acknowledgements

M.P. Coll is funded by a postdoctoral fellowship from the *Fonds de Recherche du Québec-Santé*. Geoffrey Bird is funded by an award from the Baily Thomas Charitable Trust. Essi Viding is a Royal Society Wolfson Research Merit Award holder. Claus Lamm is funded from a grant by the Austrian Science Fund (grant number P29150). We wish to thank Georgeana Bird for helpful discussion.

References

Batson, C. D. (2009). These things called empathy: eight related but distinct phenomena. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–16). The MIT Press. doi:10.7551/mitpress/9780262012973.003.0002

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, *55*(1), 19–39.

Bird, G., & Cook, R. (2013). Mixed emotions: the contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, *3*, e285. doi:10.1038/tp.2013.61

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

Brewer, R., Cook, R., Cardi, V., Treasure, J., & Bird, G. (2015). Emotion recognition deficits in eating disorders are explained by co-occurring alexithymia. *Royal Society Open Science*, *2*(1), 140382. doi:10.1098/rsos.140382

Cheng, Y., Lin, C.-P. P., Liu, H.-L. L., Hsu, Y.-Y. Y., Lim, K.-E. E., Hung, D., & Decety, J. (2007). Expertise Modulates the Perception of Pain in Others. *Current Biology: CB*, *17*(19), 1708–1713.

Coll, M.-P., Tremblay, M.-P. B., & Jackson, P. L. (2017). The effect of tDCS over the right temporo-parietal junction on pain empathy. *Neuropsychologia*, *100*, 110–119. doi:10.1016/j.neuropsychologia.2017.04.021

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science*, *24*(5), 723–732. doi:10.1177/0956797612463582

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A Review of the Concept. *Emotion Review: Journal of the International Society for Research on Emotion*, *8*(2), 144–153.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

De Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441. doi:10.1016/j.tics.2006.08.008

De Waal, F. B. M. (1996). *Good Natured* (illustrated, reprint, revised.). Harvard University Press.

De Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*, *59*, 279–300. doi:10.1146/annurev.psych.59.103006.093625

De Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nature Reviews. Neuroscience*, *18*(8), 498–509. doi:10.1038/nrn.2017.72

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. doi:10.1177/1534582304267187

Decety, J., & Meyer, M. (2008). From emotion resonance to empathic understanding: a social developmental neuroscience account. *Development and Psychopathology*, *20*(4), 1053–1080. doi:10.1017/S0954579408000503

Decety, J., Yang, C.-Y., & Cheng, Y. (2010). Physicians down-regulate their pain empathy response: an event-related brain potential study. *Neuroimage*, *50*(4), 1676–1682. doi:10.1016/j.neuroimage.2010.01.025

Devlin, H. C., Zaki, J., Ong, D. C., & Gruber, J. (2016). Tracking the Emotional Highs but Missing the Lows: Hypomania Risk is Associated With Positively Biased Empathic Inference. *Cognitive Therapy and Research*, *40*(1), 72–79.

Eisenberg, N, Hofer, C., & Vaughan, J. (2007). Effortful control and its socioemotional consequences. *Handbook of Emotion …*.

Eisenberg, N., & Sulik, M. J. (2012). Is Self-Other Overlap the Key to Understanding Empathy? *Emotion Review*, *4*(1), 34–35. doi:10.1177/1754073911421381

Fan, Y.T., Chen, C., Chen, S.C., Decety, J. and Cheng, Y., 2013. Empathic arousal and social understanding in individuals with autism: evidence from fMRI and ERP measurements. Social Cognitive and Affective Neuroscience, 9(8), pp.1203-1213.

Gu, X., Liu, X., Guise, K. G., Naidich, T. P., Hof, P. R., & Fan, J. (2010). Functional dissociation of the frontoinsular and anterior cingulate cortices in empathy for pain. *The Journal of Neuroscience*, *30*(10), 3739–3744. doi:10.1523/JNEUROSCI.4844-09.2010

Hadjikhani, N., Zürcher, N. R., Rogier, O., Hippolyte, L., Lemonnier, E., Ruest, T., ... & Helles, A. (2014). Emotional contagion for pain is intact in autism spectrum disorders. Translational psychiatry, 4(1), e343.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: in(ter)dependence of sociocognitive processes. *Annual Review of Psychology*, *68*, 243–267. doi:10.1146/annurev-psych-010416-044046

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Sciences*, *2*(3), 96–99. Retrieved from http://journals.sagepub.com/doi/pdf/10.1111/1467-8721.ep10770953

Heaton, P., Reichenbacher, L., Sauter, D., Allen, R., Scott, S., & Hill, E. (2012). Measuring the effects of alexithymia on perception of emotional vocalizations in autistic spectrum disorder and typical development. *Psychological Medicine*, *42*(11), 2453–2459. doi:10.1017/S0033291712000621

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*(4), 730–742. doi:10.1037/0022-3514.59.4.730

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, *24*(3), 771–779. doi:10.1016/j.neuroimage.2004.09.006

Keysers, C., & Gazzola, V. (2014). Dissociating the ability and propensity for empathy. *Trends in Cognitive Sciences*, *18*(4), 163–166. doi:10.1016/j.tics.2013.12.011

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self-other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1686), 20150083. doi:10.1098/rstb.2015.0083

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

Levenson, R. W., & Ruef, A. M. (1992). Empathy: a physiological substrate. *Journal of Personality and Social Psychology*, *63*(2), 234–246.

Lockwood, P. L., Bird, G., Bridge, M., & Viding, E. (2013). Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. Frontiers in human neuroscience, 7.

Meyza, K. Z., Bartal, I. B.-A., Monfils, M. H., Panksepp, J. B., & Knapska, E. (2017). The roots of empathy: Through the lens of rodent models. *Neuroscience and Biobehavioral Reviews*, *76*(Pt B), 216–234. doi:10.1016/j.neubiorev.2016.10.028

Nemiah, J. C., Freyberger, H., Sifneos, P. E., & Others. (1976). Alexithymia: a view of the psychosomatic process. *Modern Trends in Psychosomatic Medicine*, *3*, 430–439.

Panksepp, J. B., & Lahvis, G. P. (2011). Rodent empathy and affective neuroscience. *Neuroscience and Biobehavioral Reviews*, *35*(9), 1864–1875. doi:10.1016/j.neubiorev.2011.05.013

Panksepp, J., & Panksepp, J. B. (2013). Toward a cross-species understanding of empathy. *Trends in Neurosciences*, *36*(8), 489–496. doi:10.1016/j.tins.2013.04.009

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–20; discussion 20. doi:10.1017/S0140525X02000018

Prkachin, K. M., Solomon, P. E., & Ross, J. (2007). Underestimation of pain by health-care providers: towards a model of the process of inferring pain in others. *The Canadian Journal of Nursing Research - Revue Canadienne de Recherche En Sciences Infirmieres*, *39*(2), 88–106.

Rütgen, M., Seidel, E.-M., Riečanský, I., & Lamm, C. (2015). Reduction of Empathy for Pain by Placebo Analgesia Suggests Functional Equivalence of Empathy and First-Hand Emotion Experience. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(23), 8938–8947.

Rütgen, M., Seidel, E.-M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C., … Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), E5638–46. doi:10.1073/pnas.1511269112

Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: comparisons and convergence. Trends in Cognitive Sciences.

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, *17*(1), 18–24. doi:10.1177/1073858410379268

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain, 132(3), 617-627.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*, 81–96. doi:10.1111/j.1749-6632.2009.04418.x

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

Tell, D., Davidson, D., & Camras, L. A. (2014). Recognition of emotion from facial expressions with direct or averted eye gaze and varying expression intensities in children with autism disorder and typically developing children. *Autism research and treatment,* 2014.

Wang, J., Wang, Y., Hu, Z., & Li, X. (2014). Transcranial direct current stimulation of the dorsolateral prefrontal cortex increased pain empathy. *Neuroscience*, *281*, 202–207. doi:10.1016/j.neuroscience.2014.09.044

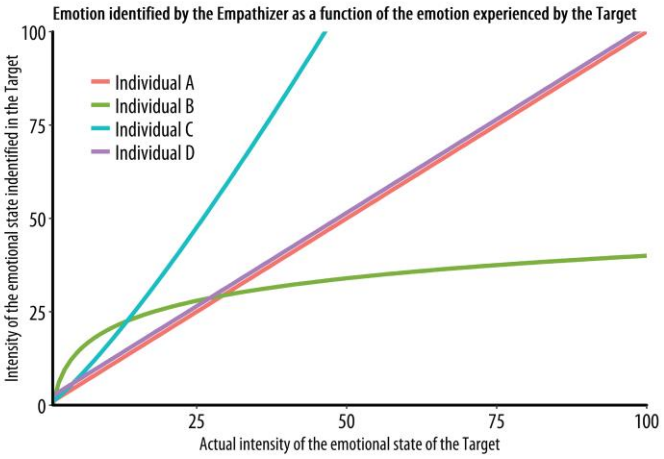
Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: the interpersonal nature of empathic accuracy. *Psychological Science*, *19*(4), 399–404. doi:10.1111/j.1467-9280.2008.02099.x

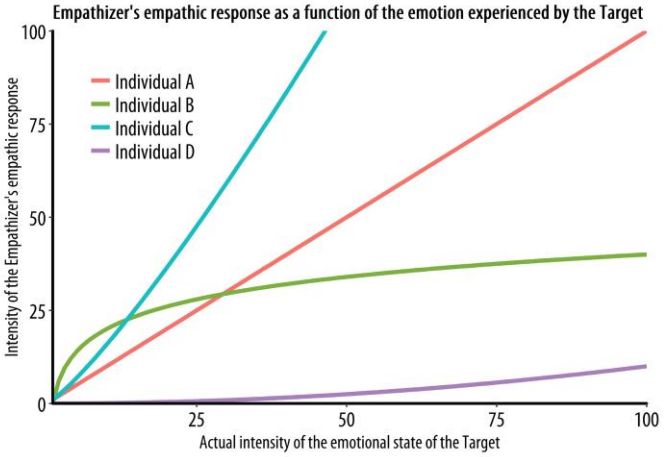
Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience*, *15*(5), 675–680. doi:10.1038/nn.3085

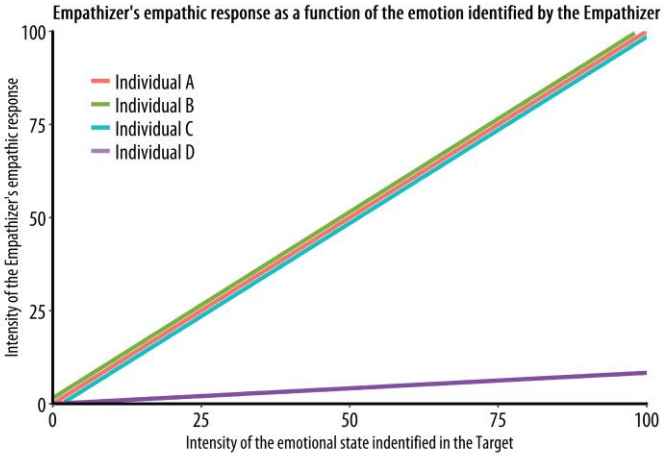
627

628 **Figures**

629







630

631

632

633

Figure 1. Theoretical graphs illustrating the relationship between the emotion experienced by the Target and the emotion identified by the Empathizer (top panel), the relationship between the Empathizer's empathic response and the emotion experienced by the Target (middle panel) and the relationship between the same empathic response and the degree of emotion identified in the Target by the Empathizer (bottom panel) for four different individuals. Individual A shows perfectly accurate emotion identification (top panel) and a degree of affect sharing which could be described using the equation for a straight line in the following manner [Emotion Elicited = 1 x Emotion Identified + 0], producing an empathic response that is perfectly concordant with the emotion they identify in the Target (bottom panel). They also meet the standard definition of empathy as they are in the same state as the Target (middle panel). Individuals B and C have less accurate emotion identification ability (top panel), tending to overestimate or underestimate the intensity of the Target's emotional state, respectively. However, they both show an empathic response which is concordant with the emotion they identify in the Target (bottom panel), and so would be judged to have the same degree of affect sharing, described using the same equation, as Individual A. Note that Individuals B and C would not meet the standard definition of empathy as their state does not match that of the Target (middle panel). Individual D is excellent at identifying the state of the Target (top panel). However, this individual's degree of affect sharing is significantly less than that of Individuals A, B and C – the slope of the line describing their empathic response in response in response to emotion identified in the Target is significantly less than that of the other individuals Emotion Elicited = 0.1 x Emotion Identified + 0] (bottom panel). This individual would also not meet the standard definition of empathy as their state does not match that of the Target (middle panel), but the source of their lack of empathy is very different to that of Individuals B and C.

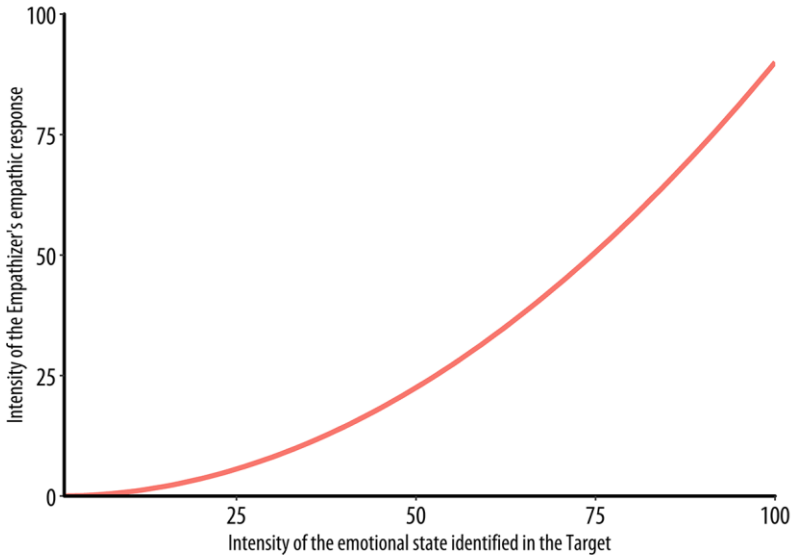


Figure 2. Graphs illustrating the relationship between the intensity of the empathic response elicited in the Empathizer as a function of the intensity of the emotional state identified in the Target. We characterise this relationship as affect sharing, and three measures may be of interest: 1) the intensity of Target emotion at which the Empathizer's empathic response is non-zero, 2) the slope of the function (indicating the degree to which changes in the Target's state prompt changes in the Empathizer's state), and 3) the shape of the function (of particular interest would be an exponential function which may indicate a dynamic interaction between empathy and emotion attribution at higher intensities of Target emotion).

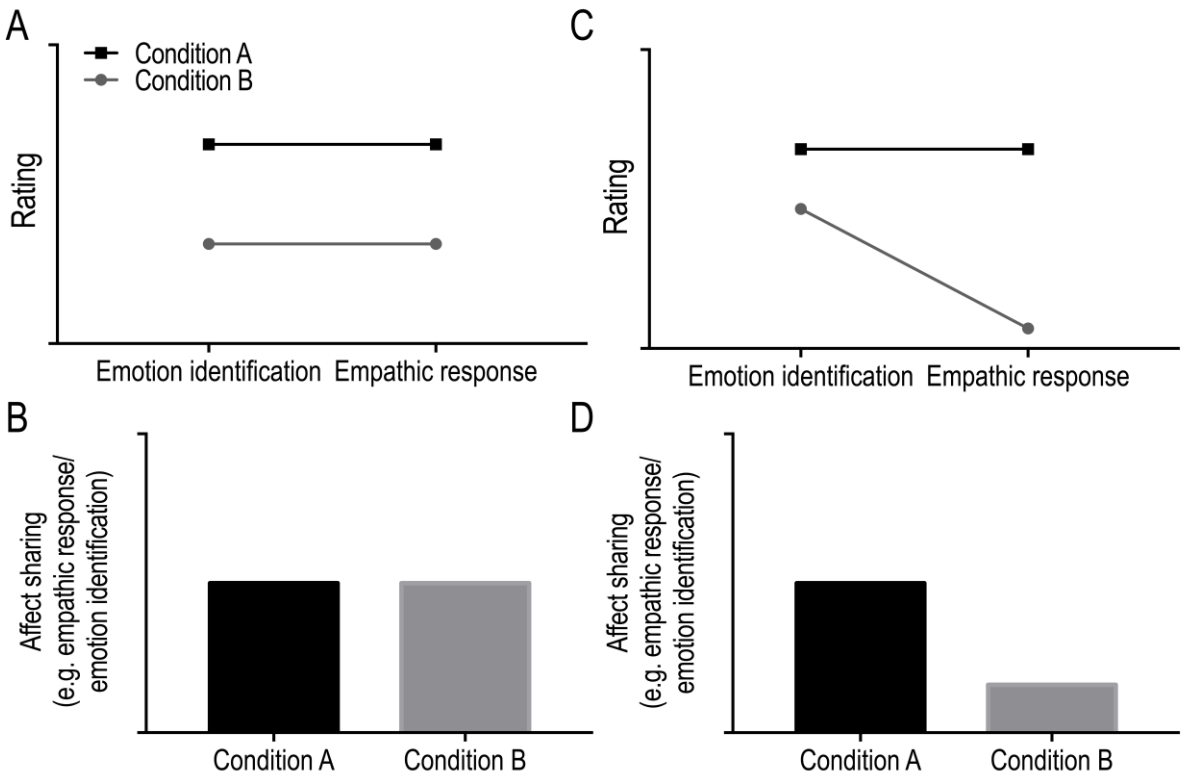


Figure 3. Hypothetical illustrations of the relationship between emotion identification, affect sharing and the empathic response. A manipulation leading to a decrease in emotion identification with a corresponding decrease in the empathic response is a sign of an absence of a change in affect sharing (A and B). The hypothetical manipulation illustrated in C and D illustrates a decrease in emotion identification, together with a greater decrease in empathic response, indicating a decrease in emotion identification and affect sharing.

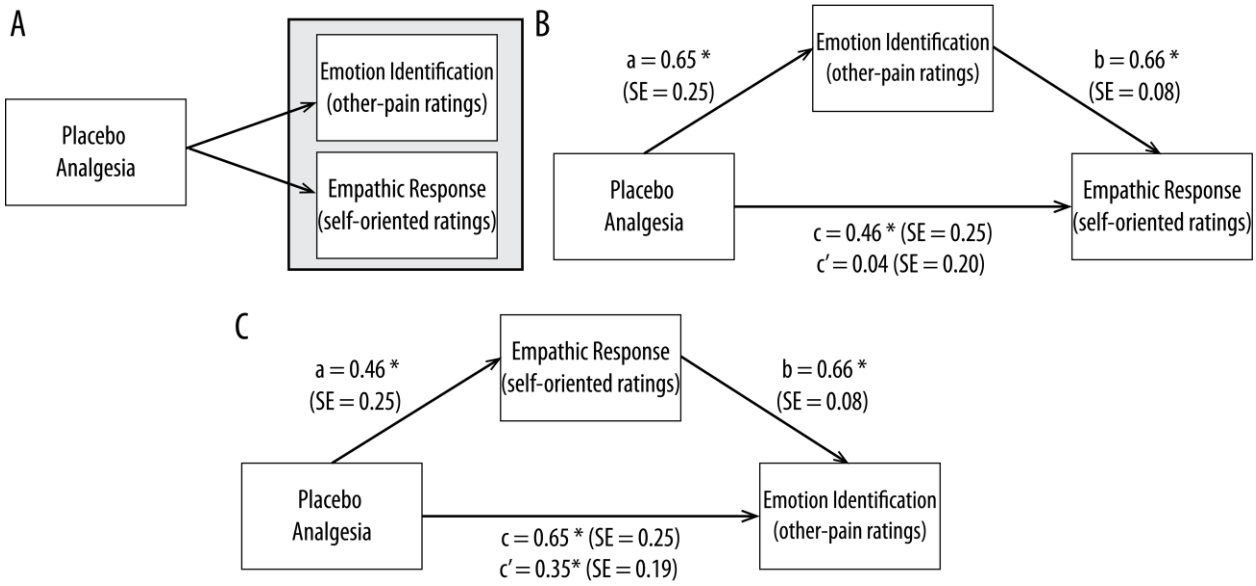


Figure 4. Three possible models of the relationships between the placebo analgesia manipulation, empathic response and emotion identification as reported in Rütgen et al. (2015). In A, both empathic response and emotion identification are independently influenced by the placebo analgesia manipulation. A mediation analysis of the Rütgen et al. data did not support this model, instead, as presented in B, the data demonstrate that the change in empathic response was fully mediated by changes in emotion identification. C shows that the empathic response also explains some, but not all, of the variance of the experimental effect on emotion identification. *$p < 0.05$, one-tailed.