



# Mirror neurons and their function in cognitively understood empathy



Antonella Corradini\*, Alessandro Antonietti

Department of Psychology, Catholic University of the Sacred Heart, Largo Gemelli 1, 20123 Milano, Italy

## ARTICLE INFO

### Article history:

Available online 11 April 2013

### Keywords:

Mirror neurons  
Empathy  
Reenactive empathy  
Rational explanation  
Social cognition  
Mindreading  
Theory–theory  
Simulation theory  
Emotion  
Intention understanding

## ABSTRACT

The current renewal of interest in empathy is closely connected to the recent neurobiological discovery of mirror neurons. Although the concept of empathy has been widely deployed, we shall focus upon one main psychological function it serves: enabling us to understand other peoples' intentions. In this essay we will draw on neuroscientific, psychological, and philosophical literature in order to investigate the relationships between mirror neurons and empathy as to intention understanding. Firstly, it will be explored whether mirror neurons are the neural basis of our empathic capacities: a vast array of empirical results appears to confirm this hypothesis. Secondly, the higher level capacity of reenactive empathy will be examined and the question will be addressed whether philosophical analysis alone is able to provide a foundation for this more abstract level of empathy. The conclusion will be drawn that both empirical evidence and philosophical analysis can jointly contribute to the clarification of the concept of empathy.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

The mirror neuron system (MNS) has been recently proposed as the biological basis of social cognition (e.g., Pineda, 2009). This encompasses a broad range of phenomena, which includes, among others, empathy (Gallese, Gernsbacher, Heyes, Hickok, & Iacoboni, 2011, Question 6). The term “empathy” is used to denote different phenomena (Roganti & Ricci Bitti, 2012). It is sometimes deployed to refer to simple forms of behavioural sharing, as occurs in emotional contagion: when a person is performing an action which is usually associated with the experience of a given emotion, another displays the same behaviour (de Vignemont & Singer, 2006). This is the case of a baby who begins crying because another baby close to her is crying or the case of laughter which spreads in a group even though people are not aware of why the others are laughing. On the other hand, empathy can be conceived of as a mainly cognitive phenomenon, which allows us to figure out the propositional attitudes that are at the basis of another's deciding, planning, and acting. Emotional aspects are not excluded, but they play a minor role in the empathic process.

In the light of this special issue's topic, empathy can be conceived of as a person's capacity to understand what others intend to do by experiencing the sensations, emotions, feelings, thoughts, beliefs, and desires which the other is experiencing (or has previously experienced). The assumption is that, if we experience the mental states of a fellow person, we can understand her reasons for her acting in a given way, and thus understand the intentions underlying her behaviour. For instance, if we realise, by watching Tom, that he has been offended by Dick and that he is now becoming angrier and angrier as a consequence of such an offence, we can understand why Tom behaves aggressively towards Dick. In turn, the comprehension of another's mental states is based, beside verbal communication, on the overt behaviour displayed by her (Avenanti & Aglioti,

\* Corresponding author. Fax: +39 02 72342280.

E-mail addresses: [antonella.corradini@unicatt.it](mailto:antonella.corradini@unicatt.it) (A. Corradini), [alessandro.antonietti@unicatt.it](mailto:alessandro.antonietti@unicatt.it) (A. Antonietti).

2006), thus the observation of others' bodily signals can be an important source of intention ascription. For instance, as suggested by Wolpert, Doya, and Kawato (2003), facial expressions – one of the main body signals we use to communicate, intentionally or incidentally, our emotional state to the others – can be seen as actions aimed at revealing the subject's intentions.

The function of empathy in understanding others' intentions can be analysed both from a scientific and a philosophical point of view. The aim of this essay is to address this topic from both perspectives. From the viewpoint of scientific inquiry, the distinction between different forms of detecting others' intentions is taken into account by referring to recent psychological literature. This will allow us to identify the specific form of empathy which is allegedly associated with MNS. Then, empirical data supporting the role of MNS in empathy will be shortly reviewed. The first section of the essay will end with some critical remarks about the need for conceptual clarification when appealing to MNS to ground empathy. These comments, by stressing the necessity of a fine-tuned analysis of such conceptual issues, will build the bridge to the philosophical section. This mainly focuses on whether reenactive empathy, that is to say cognitively understood empathy, can be conceived of as a genuine epistemic capacity, able to justify rational explanation. After a short introduction to the topic of empathy in contemporary social sciences, part 3.2. will be devoted to a defence of the soundness of rational explanation against criticisms raised by Hempel and other authors belonging to the empiricist tradition. In part 3.3, then, two arguments will be subjected to scrutiny, whose aim is to show that only reenactive empathy is able to ensure the validity of rational explanation. The upshot will be that neither argument proves to be conclusive. This result, however, does not definitively rule out empathy as an original kind of knowledge, since empirical evidence based on mirror neurons might offer some support to this epistemological thesis, in particular if basic kinds of empathy are taken into consideration.

## 2. Empathy and MNS from the point of view of psychology and neuroscience

### 2.1. *Mirroring and mentalising mechanisms underlying empathy*

Empathy is a complex phenomenon involving different aspects and dimensions. In fact, the understanding of others' intentions through the experience of their mental states may be underwritten by different processes. On the one hand, as shown by the example reported in the previous section, we can immediately understand the reasons for Tom's aggressive behaviour on the basis of the perception of his face and/or the tone of his voice. We establish a direct connection between what Tom looks like (in terms of bodily appearance and bodily movements), his mental states, and his acts. On the other hand, we can understand Tom's intentions by integrating the perceptual information Tom provides us with and some inferences based on contextual cues (for instance, the presence of other people on the scene who are mocking him), specific notions we have about Tom (for instance, remembering that Tom is a choleric guy), and abstract concepts (for instance, our conviction that an offended man should always take revenge).

In the fields of psychology and the neurosciences some distinctions have been drawn in the attempt to clarify the mechanisms underlying the understanding of others' intentions. A relevant starting point may be the distinction which has been made, under different concepts and linguistic labels, between a system which allows human beings to comprehend immediately others' intentions and a system which allows humans to reach such an outcome through an inferential process which implies the mediating role of some forms of reasoning.

This distinction relies on a more fundamental distinction which has been recurrently proposed by different authors in recent years in the domain of thinking and decision-making processes (Sloman, 1996), namely, the distinction between the so-called System 1 and System 2. System 1 (Stanovich & West, 2000) – also labelled as intuitive (Pretz, 2008), experiential (Slovic, Finucane, Peters, & MacGregor, 2002), tacit (Hogarth, 2001), impression-based (Kahneman, 2003) – is fast and action-oriented; it is activated unintentionally and its functioning is rigid and partially behind the control of the individual. Usually it operates effortlessly on the basis of associations. System 2 (Stanovich & West, 2000) – also called analytical (Slovic et al., 2002), deliberative (Hogarth, 2001), judgment-based (Kahneman, 2003), rational (Epstein, 1994) – operates slowly, intentionally, and flexibly, predominantly on the basis of abstract representations and logical rules. Usually it is not emotionally charged. The functioning of System 2 may fail to be optimal because of the excessive cognitive load it requires, its slowness, and the large amount of effort its activation needs.

In this vein, with specific reference to social cognition, Bohl and van den Bos (2012) proposed the distinction between *Type 1* and *Type 2* process. The former is fast, efficient, stimulus-driven, and lacks flexibility. The latter is slow, involves a high cognitive load and elaboration, is flexible and accessible to consciousness. With a more specific focus on processes involved in understanding other people's intentions, Waytz and Mitchell (2011) distinguished between *mirroring* and *self-projection* mechanisms. The first mechanism enables us to understand other people by experiencing vicariously their mental states: thanks to such a mechanism, the others' mental states are mirrored in our mind. Through the second mechanism we project our mental states onto the situation of another individual, so to infer her mental states. According to the authors, the two mechanisms involve different degrees of immediacy in others' understanding. Mirroring is a sort of on-line process which allows us to resonate immediately according to what another person is experiencing; self-projection, by contrast, implies imagining off-line what we should experience if we were in the other's shoes and then attributing such an experience to her.

The distinction between mirroring and self-projection overlaps partially the distinction between *mirroring* and *mentalising* (Chiavarino, Apperly, & Humphreys, 2012). The mirroring system responds to observation of others' acts and seems to

code their goals immediately by establishing purely behavioural relations between the perceptual appearance of the actor and her corresponding intentions. Mentalising instead requires inferences about the mental states which are at the basis of the behavioural relations. More precisely, the second system has two subcomponents: a representational one, which serves the task to represent the actor's intention as a mental state (but not as a behavioural relation), and a conceptual component "representing the semantic and logical properties of intentions, abstractly reasoning over these properties, and relating them to other mental states" (Chiavarino et al., 2012, p. 286).

## 2.2. Mirror neurons and empathy: empirical data

The neural circuits constituting MNS have been proposed as the best candidate for the biological basis of empathy, which is to be thought of as the expression of the mirroring process. In fact, MNS has been invoked as a putative interpretation of empathy and some experimental findings have been taken as evidence supporting the involvement of MNS in empathy (Gallese, 2001, 2003; Iacoboni, 2009; Preston & de Waal, 2002).

First of all, it is documented that humans, when watching people showing facial expressions corresponding to well-defined emotions, covertly activate the same muscles which are involved in the creation of those expressions (Dimberg, Thunberg, & Elmehed, 2000). Moreover, if people are prevented from automatically imitating the muscle contractions of the faces they are exposed to (for instance, by compelling them to keep a pencil with the teeth transversal to the mouth), they become less able to detect the emotional expression of the observed faces (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). This experimental finding supports analogous results observed in patients affected by the Moebius syndrome, which impedes them to move their facial muscles: as a consequence of such an impairment, these patients fail to recognise the emotions expressed by others (Cole, 2001). Finally, it is worth noting that the same cortical areas are activated when people observe and imitate faces expressing emotions (Leslie, Johnson-Frey, & Grafton, 2004). Hence, it is proved that, in emotion recognition, observation and action are linked together, as in the case of the functional actions directed at manipulating things, which have been the main topic of investigation in MNS field.

These findings, however, only concern emotion recognition, which is not empathy, but rather its precursor or precondition. Further empirical evidence is required. Indeed, other studies showed that the link between observation and perception also regards empathy. For instance, if individuals are paired with a confederate who imitates their postures, gestures, and body movements during the execution of a joint task, they perceive the confederate as more agreeable than controls paired to a non-imitating confederate do (Chartrand & Bargh, 1999). In addition, individuals who spontaneously imitated the behaviour of the confederate scored higher on an empathy scale subsequently, showing a positive relation between the frequency of imitative behaviours and the empathy rates (Chartrand & Bargh, 1999).

Two emotional reactions have been often investigated in the attempt to prove the involvement of MNS in empathy: pain and disgust. As to pain, Avenanti, Buetti, Galati, and Aglioti (2005) recorded the excitability of the muscle of the hand which generates an approaching movement toward a noxious stimulus (a needle): when people looked at a video showing other people whose hand was penetrated by a needle in the same point, the excitability of the muscle decreased (as if they were trying to move away the hand from the needle); in addition, the reduction of the excitability of the muscle was proportional to the estimated level of pain the subjects attributed to other people when their hand was penetrated by the needle (see also Avenanti, Minio-Paluello, Sforza, & Aglioti, 2009; Valeriani et al., 2008).

As far as the brain counterparts of pain experience are concerned, it was showed that neurons in the anterior cingulate cortex responding to painful stimuli applied to the subject's hand also fired when the subject observed another person being stimulated by the same noxious stimuli (Hutchinson, Davis, Lozano, Tasker, & Dostrovsky, 1999). Anterior cingulate cortex, together with some regions of the insula, was also activated by observing relatives who were not currently exposed to painful stimuli, but would be stimulated painfully later (Singer et al., 2004). Hence, not only the direct observation of suffering people, but also the prefiguration of a future pain affecting others activate the brain areas corresponding to the actual experience of pain in first person.

The same message is provided by studies concerning the neural counterparts of disgust. It has been proved that the same brain structure (the insula, in this case), which is active when the individual experiences disgust personally, is activated even when the individual looks at faces expressing disgust and that the intensity of such an activation is proportional to the level of disgust expressed by the face (Phillips et al., 1997). The evidence was later supported by recording the activity of neurons in the anterior part of the insula through electrodes implanted in the brain of epileptic patients (Krolak-Salmon et al., 2003). A clear proof that the same neural counterparts are involved in experiencing disgust and observing other people experiencing that emotion was provided by Wicker et al. (2003) in a fMRI study where the same participants were both exposed to disgusting odours and to pictures of persons smelling the same odours.

The impairment in experiencing negative emotions is associated with the impairment of recognising similar emotions in other people. In fact, a case was reported of a patient with brain lesions in the putamen and in the insula who failed to subjectively experience disgust (and, as a consequence, to react to disgusting situations appropriately) and also was not able to detect disgust in other people by observing their facial expressions or by listening to non-verbal sounds which they produced, as well as to the prosodic aspects of their speech (Calder, Keane, Manes, Antoun, & Young, 2000). A similar case was successively reported by Adolphs, Tranel, and Damasio (2003).

When trying to find evidence that MNS is specifically involved in empathy, we can point to the fact that the activation of brain areas included in MNS has been recorded in participants both when they were simply looking at actors showing facial

expressions whose emotional meaning corresponded to that of the story they were telling (Decety & Chaminade, 2003) and when they were asked to identify the emotional states of actors by observing their body postures, gestures, and facial expressions (Lawrence et al., 2006). Further support came from the experiment executed by Schulte-Rüther, Markowitsch, Fink, and Piefke (2007): mirror-neuron mechanisms were activated when participants, exposed to facial expressions, had to identify both the emotions concurrently experienced by themselves and the emotions expressed by the others' faces.

The involvement of MNS in empathy is also proved by correlational studies. Two investigations demonstrated that the level of emotional empathy developed by the participants was correlated to the intensity of the activity of premotor areas, presumably containing mirror neurons, when the participants were asked to look at other people carrying out the act of grasping with different intentions, as suggested by contextual hints (Kaplan & Iacoboni, 2006) or to listen to sounds produced by human actions (Gazzola, Aziz-Zadeh, & Keysers, 2006). More specifically, participants who showed higher activation of brain areas involved in MNS when looking at facial expressions by focussing on their emotional valence (Schulte-Rüther et al., 2007) obtained high scores on empathy scales.

MNS, together with other brain structures such as the limbic system and the insula, constitutes a large neural circuitry which has been proved to be activated by both the execution, through imitation, and observation of facial expressions associated to emotional experiences (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Iacoboni & Lenzi, 2002). The association between MNS and both the subjective experience of emotions and the detection of the same emotions in others through the observation of their behaviour is further supported by a fMRI study which showed that the activation of MNS in preadolescents while observing and imitating emotional facial expressions is positively correlated with the level of empathic skills (Pfeiffer, Iacoboni, Mazziotta, & Dapretto, 2008). An additional support is provided by clinical studies carried out with people affected by autism. On the one hand these patients – who are impaired in recognising emotions from others' facial expressions and to imitate such expressions – fail to show the usual reactions when looking at other people being affected by painful stimuli (Minio-Paluello, Baron-Cohen, Avenanti, Walsh, & Aglioti, 2009). On the other hand people with autism show deficits in MNS functioning and their level of activity of MNS is reduced in correspondence with the level of severity of the pathology (Dapretto et al., 2006).

### 2.3. Mirror neurons and empathy: conceptual problems

One of the main messages which are associated to the findings concerning the involvement of MNS in the understanding of others' mental states, including intentions, is that such an understanding does not exclusively depend on linguistic and mentalistic processes (Gallese, 2001, p. 34). On the contrary, intentions are embodied. Such an embodiment is shared both by the actor and the observer and relies on the motor schema of action. When the motor schema of the actor matches a motor schema in the repertoire of the observer, the intended meaning of the action is detected (Gallese, 2001, p. 36). If this general framework is applied to empathy, the consequence is that empathy is grounded in the experience of the lived body: others are conceived “not as bodies endowed with a mind but as persons like us” (Gallese, 2001, p. 43). In this way we can recognise why persons behave in a certain manner.

In some circumstances, the comprehension of the intentions of others' behaviour occurs predominantly on the basis of the emotions they are experiencing rather than of the functions of the actions they are performing. When this happens, empirical findings summarised above suggest that MNS is involved, either because some cerebral areas belonging to MNS are directly activated or because other brain structures, connected to the main mirror-neuron areas, are activated, such that they successively involve the proper mirror-neuron areas. In any case, the resulting outcome is that the same brain structures, which are activated when we experience the affective state the other is experiencing, are activated. This would lead the affective states of other people to *resonate* in the mind of the perceiver (Gallese, 2001, p. 38; Rizzolatti & Sinigaglia, 2006, p. 121) or, put differently, would generate in the perceiver a sort of *inner imitation* of what the other is feeling (Iacoboni, 2008, chap. 4). Another way of thinking of this is that the cerebral system of the observer would be activated *as if* she were behaving as the observed human being. This occurs because the observed behaviour is translated into a program which acts as a sort of signal (efference copy signal) which enables the *simulation* of the behaviour (Gallese, 2001, pp. 40–41). As a consequence, the other's behaviour is *modelled* as an action thanks to the behavioural equivalence between the perceiver's and the other's actions (Gallese, 2001, p. 39). A first critical remark is that further clarification of the mental process supported by MNS during an empathic relation is required. Resonance, inner imitation, simulation, and modelling are different processes and the authors claiming that MNS grounds empathy should be more explicit and precise about the psychological counterparts of the corresponding cerebral activations.

Whatever these processes may be which are supported by MNS and lead to empathy, authors maintaining that MNS is involved in empathy generally agree that intention understanding does not involve any form of abstract thought. To put it in the authors' words, it is “non-predicative” (Gallese, 2001, p. 44), “without verbal mediation” (Rizzolatti & Sinigaglia, 2006, p. 120), “without the need of theorising” (Gallese, 2001, p. 41), “without propositional attitudes” (Gallese, 2001, p. 41), “non-inferential” (Gallese, 2001, p. 44; Rizzolatti & Sinigaglia, 2006, p. 174), “without any knowledge operation” (Rizzolatti & Sinigaglia, 2006, p. 127), “not needing cognitive processes” (Rizzolatti & Sinigaglia, 2006, p. 174), “pre-reflective” (Iacoboni, 2009, p. 666). In these authors' view, MNS leads us to comprehend others' experience in the absence of any conceptual representation and inference. Now, how should this form of understanding be conceived? This is a list of the adjectives which are attributed to it: “direct” (Gallese, 2001, p. 41), “immediate” (Gallese, 2001, p. 41; Rizzolatti & Sinigaglia, 2006, p. 127), “effortless” (Iacoboni, 2009, p. 666), “automatic” (Gallese, 2001, p. 41; Iacoboni, 2009, p. 666), “implicit” (Gal-

lese, 2001, p. 41), “unconscious” (Gallese, 2001, p. 41), “subpersonal” (Gallese, 2001, p. 42 and 46). Here, too more precision seems to be required (Debes, 2009). In fact these attributes have different meanings and do not implicate one another. For instance, pure knowledge operations and cognitive processes, with no form of embodiment, can also be immediate and effortless, if adequately trained. Also the meaning of “automatic” and “unconscious” should be specified. A process can be automatic by its very nature or because it has become such after having been carried out for a long time with effort and the labour of reasoning. The same is true of the unconscious character of intention understanding: is it a process which has become unconscious as a consequence of its automatization or because it has always been unconscious? In other words: the process might be conscious (and involving effort) when the individual is trying to learn to carry it out, but it becomes unconscious (and effortless) when she had learnt to master it. In addition: does the unconscious character of the process make reference to how the process develops or to the outcome of the process? We can be unaware of how we compute the sum  $5 + 2$ , but we are aware of the output of the process (and also of the fact that we are computing the sum). Also the claim that intention understanding supported by MNS through empathy fails to involve knowledge and cognitive mechanisms can be questioned. As noted by Roganti and Ricci Bitti (2012, pp. 583–584), appraisal processes are always implied in emotion comprehension, and thus an interpretative component can never be discarded, otherwise only a form of emotional synchronisation or synthonisation, but not a real understanding, occurs. Thus, the specific forms of cognition which should be excluded by the kind of empathy supported by MNS have to be clarified, since it has been proved that other cortical regions, beside MNS, are involved in cognitive manifestations of empathy (Shamay-Tsoory, Aharon-Peretz, & Perry, 2008).

In conclusion, it appears that a more fine-grained analysis of the features of the empathic relation supported by MNR is needed. To this end, this issue has to be addressed from the philosophical perspective, which we turn to now.

### 3. From mirror neurons to reenactive empathy

#### 3.1. Empathy as reenactive empathy

As the first part of this essay has shown, the renewed, current interest in empathy is strictly related to empirical research in the fields of neurobiology and psychology. In particular, the discovery of MNS in monkeys has given new impulse to the scientific treatment of empathy. However, the notion of empathy has a long philosophical tradition, characterised by many ramifications and several divergent approaches (for an informed reconstruction of the history of empathy see Stüber, 2006, Introduction, and 2008). As far as philosophy of the social sciences is concerned, the most influential twentieth century supporter of empathy has been the philosopher of history Robert Collingwood (1949) who, against explanatory monism, maintained that explanation in history requires an essential empathic component. In fact, we cannot explain the behaviour of a historical character without re-enacting her intentions, beliefs, desires and choices. Yet, the role of reenactive empathy has not always been positively evaluated within the philosophy of the social sciences, partly because it introduces a sharp dualism between natural and social sciences, partly because it appears to represent a capitulation to any sort of subjectivism and arbitrariness (see Popper’s criticism of the epistemological role of empathy in Popper, 1972, 4.12). In recent years, however, authors such as Jane Heal and Karsten Stüber have revived the fortunes of empathy and have argued in favour of a strict correlation between rational explanation and empathy as a fundamental epistemic capacity. Most part of what follows is a discussion about the theses put forward by both authors.

#### 3.2. Rational explanation

As is well known, there are two main ways of conceiving explanation in folk-psychology. The first is theory–theory and the second is simulation theory.

According to theory–theory, human actions are explained on the basis of the classical Hempelian method, which, although imperfect in a number of ways, nonetheless maintains its fundamental validity. What is essential in this method is the presence within the *explanans* of empirical laws having the form of universal conditionals. From the point of view of theory–theory supporters, then, action explanation is an empirical theory, which explains agents’ actions through empirical laws, just as any empirical theory explains the behaviour of certain objects. The laws may be different from some other empirical theories, since in folk-psychology they often are probabilistic or *ceteris paribus* laws; nonetheless, the explanatory structure is the same. By contrast, according to simulation theory, an agent’s behaviour is explained through simulation of the reasons, beliefs and desires which move the agent to action.

An example of the theory–theory paradigm consists in the last of the three inferences involved in a successfully performed false belief task.

“Predicting where Maxi will look for the chocolate.

- i. Maxi wants to eat the chocolate, and he believes that the chocolate is in cupboard, and he believes that looking for the chocolate in cupboard is a means of satisfying one’s desire of eating it.
- ii. *Central action principle*: If somebody desires *x* and believes that *A*-ing is a means of achieving *x*, then, *ceteris paribus*, he will do *A*.
- iii. Max will look for the chocolate in cupboard.” (Stüber, 2006, pp. 109–110).



The issue at stake is to clarify the nature of Central Action Principle (also named by Kim (1984, 1998) Desire/Belief/Action Principle – DBA), which in the inferential scheme fulfils the function of a general law. For theory–theory supporters this principle is an empirical law expressing a kind of nomological necessity and its presence is necessary. The notion of nomological necessity is, in fact, essential in order to define the relation of causal explanation.

Critics of the theory–theory paradigm, however, cast doubts on the empirical nature of laws like DBA. What they question is not their general character; rather, critics point at the fact that they do not express a nomological necessity but a logical–analytical necessity, which sometimes is defined as a “rational” necessity (Dray, 1957). In other words, principles of this kind are principles that define the notion of rationality. They are the axioms of practical rationality, formally encoded in practical syllogisms (or practical inferences) (Anscombe, 1957; von Wright, 1971).

Hempel’s criticism of rational explanation is well known (Hempel, 1965, chap. 5). If the DBA principle is a rationality axiom, it cannot be the object of empirical confirmation and, thus, it cannot be considered as a principle endowed with explanatory capacity. To avoid this consequence it is necessary to modify DBA to the following principle, DBA\*.

*Central action principle revised:* If somebody desires *x* and believes that *A*-ing is a means of achieving *x* and *she is a rational agent*, then, *ceteris paribus*, she will do *A*.

However, this move is fatal to rational explanation, since a rationality clause cannot be included among the particular facts of a law that claims to contribute to the definition of rationality itself. Indeed, to obtain confirmation of a law like DBA\* we must be in the business of establishing the truth of the antecedent, including the rationality clause. But, how is it possible to establish an agent’s rationality, if we need the principle DBA\* itself to define the rationality notion?

The contrast between rational and nomological explanation seems to be so stark as to only allow two possibilities. Either we maintain that explanation should be nomological in the human sciences just as in the physical sciences, which almost necessarily leads to a naturalistic re-interpretation of folk-psychology. Or we give up the claim that actions are explainable as human actions and fall back on the less ambitious idea that they can only be the object of understanding; that is to say they are behaviours that we can interpret in the light of an agent’s subjectivity but that have their cause elsewhere. Is there a way out of this dilemma?

Borrowing from Stüber (2003) we assume that principles like DBA can be conceived of both as analytical principles and as empirical generalisations. If they are understood as analytical principles, they express a necessity of a conceptual kind and have, on top of that, a normative meaning, as they formalize a correct way of reasoning. An agent who does not abide by them does not reason correctly and, as a consequence, does not decide correctly. These principles, however, can also be conceived of as empirical generalisations, inasmuch as they describe the way agents “in flesh and blood” reason and take decisions. In this latter meaning, and only in this latter meaning, they are falsifiable by experience. “The distinction between understanding a general statement as the articulation of a normative standard or as the description of a regularity in behaviour points to different functions of the same statements in different contexts” (Stüber, 2003, p. 268).

But why can they *also* be conceived of as empirical generalisations? Expanding on the previous argument, we can give the following answer. Analytically understood DBA principles define the concept of rationality. As we said before, they are rationality axioms and, thus, define the way a real mind (or a mechanism like a mind) should function in order to be a mind that operates rationally. It is worth noting that, from this viewpoint, also principles belonging to scientific theories could be conceived of as axioms that define certain models of empirical reality. In this way theory confirmation would not be anything other than the confirmation of the fact that the model defined by the theory is actually instantiated in empirical reality.

Owing to this analysis of DBA, the rationality clause, which Hempel considered as a necessary condition in order to justify the explanatory character of practical argumentation, can now be put in the right place. For the previously mentioned reasons, such a clause must not be put among the other clauses of the conditional that makes the law. It must be considered, instead, as a fundamental presupposition, that is to say as a background assumption, that permits us the very use of those laws. In other words, it is the same assumption made in the practical–inferential model of rationality, understood as an explanatory model of empirical agents’ concrete actions. This assumption could be of course totally wrong, if the model did not work at all, that is to say if no set of actions did exist, which can be explained by any exemplification of the model. However, this can hardly be the case, since this would be tantamount to saying that the set of the actions to be rationally explained is empty. Instead, it is easier to falsify such a hypothesis on particular occasions, in which the action under scrutiny does not derive from true premises of the DBA principle. However, also on these particular occasions it is not the falsification of DBA as an analytical (logically correct) principle that occurs, but the falsification of the principle considered as an explanatory scheme. Thus, in the end, it is the falsification of the validity of the principle in that particular case.

Yet, at this point a problem arises. The correctness of the DBA scheme can be justified by means of a priori reasons. It is an analytical principle and, therefore, it has to be founded in a similar way as those of the formal sciences. But what about DBA as an explanatory law, that is to say as the fundamental presupposition concerning the explanatory dimension of the practical–inferential model in its application to reality?

### 3.3. *Is reenactive empathy an epistemic capacity?*

Authors such as Heal and Stüber answer this question by recourse to the empathy thesis. We are supposed to have the direct perception of the connections established by the DBA principle among our desires, our beliefs and the actions we perform. This perception is a first-person perception. It is the agent’s ego who perceives the logic of her acting. At this point

empathy enters the scene. The agent is able to put himself in others' shoes, meaning not only that she succeeds in grasping the others' point of view, but also in reenacting the same experiences which characterise others' mental and emotive processes. In this way empathy becomes the *epistemic* element able to endow schemes like DBA with an explanatory dimension: that is to say empathy guarantees the soundness of their explanatory use. As Stüber puts it: "The special status of DBA just derives from the fact that we ourselves have access to our cognitive states and reasoning from the first person perspective and have to use this ability in a projective or reenactive manner in order to understand the reasons of other agents" (Stüber, 2003, pp. 275–276; Stüber, 2006, pp. 212–213).

But why should we resuscitate reenactive empathy or co-cognition (Heal, 2003, p. 97 ff) in order to account for the soundness of the practical inferential explanation of actions? Heal (2003) and Stüber (2006, p. 152 ff) put forward two arguments in favour of the epistemic relevance of empathy. In the remainder of this essay I will expound and comment on both arguments.

#### (A) The argument from the essential contextuality of thoughts as reasons

As the argument goes, theory–theory aims to deliver a complete theory of action, including both a formal inferential scheme to allow people's thoughts to be related together until the final decision is taken and a criterion to objectively establish the premises of the inferential scheme.

But such a theory does not exist, because there is no complete formal inferential scheme and, above all, there are no criteria to objectively establish the premises of the inferential scheme, as they are dependent on the context. In fact, according to the *frame problem*, in order to establish the premises it is necessary to know what are the relevant aspects for ends choice. But to understand what are the relevant beliefs for explaining an action it is necessary to understand what are the beliefs that are meaningful to themselves as a subject, and this is not possible without reenactive empathy.

Analysis of and comment on the argument.

Three aspects of the previous argumentation should be distinguished.

- a. Theory–theory is criticised because it is unable to deliver a complete theory of action: it only partly covers the process of practical decision. In particular, it cannot solve the problem of the premises identification. Thus the theory is insufficient, since it is incomplete as regards its premises.
- b. Theory–theory is criticised because it is objective, that is, it does not account for the first person's perspective.
- c. Theory–theory is criticised because it cannot justify the attribution of a causal role to agents' desires and beliefs when these concern strictly subjective contents. But – as we learned from the contextuality argument – how is it possible to explain agents' behaviours if we do not have any access to their subjectivity, that is, if we do not succeed in understanding how and why these contents are meaningful to the agent? Such an access presupposes an original and irreducible capability to empathically identify ourselves with a subject different from us.

Now, our comment is that reenactive empathy may be needed, if it is needed at all, to provide a solution for the problem formulated at point c, but neither for the problem mentioned at point a nor for that at point b. Let us ask ourselves, in fact, what is needed to overcome incompleteness (point a).

A method is needed that is able to establish the premises, which, however, are formulated from the subject's viewpoint. They are premises which do not express states of affairs of the agent's life that can be described as causes or objective conditions of the agent's conduct. The states of affairs corresponding to such premises are not characterised by empirically detectable properties, thus we cannot grasp them without taking into account the subjective framework in which they are situated. In other words, they are states of affairs that can be described as structurally identical with the desires and beliefs from which our actions originate. Thus they presuppose the access to the first person's perspective. We can then conclude that in order to overcome the incompleteness of the theory we need to modify it so to make it able to express the first person's perspective. What is it necessary for such a goal? To say that we need reenactive empathy appears to us to be too a hasty strategy. Instead, we need to replace the Hempelian D–N scheme with the P–I scheme of practical inference. Practical inference, in fact, consists in a general scheme including assertions of the goals to be reached (B(x, goal p) as premises, beliefs on the chain of actions to perform for reaching those goals (B(x, N(p<sup>®</sup> to do q) and, as a conclusion, the action the subject decides to perform (x does q). Premises and conclusion are not connected to each other by empirical laws, but by principles belonging to practical logic (epistemic and deontic logic at first place). It is this very essential aspect of P–I that makes it capable of expressing the first person's perspective. Actually, the above outlined scheme works perfectly if we replace x with the indexical "I", a typical expression of the first person's perspective. So far, however, no kind of reenactive empathy is needed. We are dealing with the first person's perspective, which only requires the subject's capacity to grasp the nexus between the practical–inferential scheme and the world of her own desires and beliefs. To this end a form of self-perception or of self-awareness, but not of empathy is needed. Reenactive empathy could perhaps play a role at point c, when the P–I scheme is transferred to other subjects. We can in fact ask ourselves what would justify the extension of the scheme to other subjects and how would it be possible to explain their actions through a scheme like P–I if we could not accede the other's world as if it were our own world. This part of the argument from contextuality will be expanded on in the second argument from the indexicality of thoughts as reasons (Stüber, 2006, p. 161 ff).

## (B) The argument from the essential indexicality of thoughts as reasons

This argument is divided into two parts.

The first part of the argument can be formulated in the following way.

The scheme:

1.  $B(x, \text{goal } p) \ni B(x, N(p \rightarrow q)) \rightarrow x \text{ makes } q \text{ appears to me as explicative of the actions } x \text{ does only if it appears to me as explicative of the actions I do. But it is explicative of my actions only if I understand that my belief that goal } p \text{ and my belief that } N(p \rightarrow q) \text{ imply my doing } q. \text{ In other words, my reasons for my doing } q \text{ must be ego-indexed.}$
2. The second part of the argument reads that when I think of goal  $p$  and  $N(p \rightarrow q)$  as reasons for others, I have to think of them as if they were *my own* reasons.

Analysis of and comment on the argument.

- a. Our first remark is that indexicality essentially requires the first person's perspective. However, this does not imply empathy, but only self-perception or inner perception.
- b. Secondly, empathy as an ability to reenact others' thoughts is required, if at all, in the passage from my reasons to others' reasons. That is, it has to be introduced as early as in the second part of the argument from indexicality.
- c. Thirdly, it is debatable whether reenactive empathy should be understood as a *sui generis* kind of knowledge. Empathy is in fact a form of knowledge which, on the one hand, is different from empirical perception, but which, on the other, is of the very same kind as perception, that is to say it is a form of perception and not a sort of pure a priori evidence. Thus it would represent a third kind of knowledge, in addition to empirical knowledge and genuinely a priori knowledge. The question then is whether it is strictly necessary to assume this new form of knowledge or whether it could be replaced by the synergic work of both inner perception and a priori knowledge. Empathy could be understood as an intentional capacity addressed to a "something" which is different from ourselves and is conceived of as a subject rather than as an object. The knowledge we suppose to have of the other could be actually interpreted in a different manner, that is, as the result of the information we derive from our capacity to represent to ourselves the other's world and to draw from this the explanation of her behaviour. On this construal, we would be entitled to believe in the existence of others' inner worlds thanks to their capacity to explain others' behaviours to me.

#### 4. Conclusion

Heal's and Stüber's arguments in favour of the epistemic role of reenactive empathy do not appear to be conclusive. However, the way the argumentation has been developed does not preclude the possibility of exploiting empirical research to ascertain whether reenactive empathy is or is not an original kind of knowledge of others' minds. Clearly, the deployment of empirical data to deal with an epistemological problem requires a theory about the relationship between conceptual and empirical knowledge. Its function is to legitimate the contribution of the empirical data to the construction of a conceptual and a priori kind of knowledge like that supplied by philosophy. We cannot exhaustively treat this topic in the present essay, but, since MNS delivers the most influential empirical result so far about empathy, we shall deal with the issue of whether the empirical results lend some support to this philosophical account of empathy.

First, it is appropriate to stress that a priori and empirical considerations about empathy are not incompatible. In fact, the philosophical reflections provided above leave open the question about empathy. The possibility of empathy as an original kind of knowledge of other subjects' minds has not been excluded, even though doubts have been cast on the claim that a priori reasons are sufficient for reaching a positive verdict on empathy.

However, something more than mere compatibility is needed for arguing for the relevance of MNS to the empathy issue. It is necessary that neurobiology tells us what the role of mirror neurons is in the construction of intersubjectivity and if this role supports the thesis of a capacity which cannot be explained through the perception of a purely objective phenomenon like behaviour.

We believe that the answer to the first question is quite uncontroversial. Empirical evidence shows that mirror neurons play a major role in the construction of a basic kind of relationship with the other. The answer to the second question is instead more thorny. It implicitly contains a hint at an impossibility proof. A theory about mirror neurons, in fact, should provide an argument to the effect that mirror neurons possess a capacity for intersubjectivity that is not explainable through the mere elaboration of objective perceptive data. In other words, the theory about mirror neurons should be able to exclude the possibility that mirror neurons can perform their function without implying empathic capacities. As is known, proofs of impossibility (or of indispensability) are very difficult and, sometimes, not conclusive. An opponent of the empathy thesis based on mirror neurons could argue that she can explain the evidence of intersubjectivity in a different way, i.e. by means of evolutionary theory. The argument would go as follows. The promptness with which neurons react to others' mental world does not depend on any specific empathic capacity, but on the fact that this circumstance expresses in an immediate and non-reflexive way the capacity that the organism has acquired in its millenary history to represent to itself others' mental world on the basis of their behaviours. The mirroring capacity of mirror neurons should not then be explained empathically, since it can be conceived of as the result of an evolutionary application of theory-theory.



However plausible this interpretation may appear, we think that the evolutionary way of arguing is debatable. In fact, it could be employed to all of the a priori capacities. As the evolutionary theory of knowledge affirms, these too are the result of the sedimentation of the species' experiences in the individual. Thus, just as we believe that the evolutionary argument does not hold as regards the a priori in general, it seems to us quite wobbly also in the case of an empathic reading of mirror neurons.

To sum up, there is no question that philosophical reflection successfully argues for the validity of the first person's perspective, while the a priori arguments in favour of the empathy thesis remain problematic. Empirical research on MNS is surely an important clue to the validity of the empathy thesis, at least as far as empathy is conceived of as a basic capacity (on this see Corradini, 2011). Nevertheless, further work is still to be done on both sides.

## References

- Adolphs, R., Tranel, D., & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain Cognition*, 52, 61–69.
- Anscombe, G. E. M. (1957). *Intention*. Oxford, UK: Blackwell.
- Avenanti, A., Buetti, D., Galati, G., & Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Neuroscience*, 8, 955–960.
- Avenanti, A., Minio-Paluello, I., Sforza, A., & Aglioti, S. M. (2009). Freezing or escaping? Opposite modulations of empathic reactivity to the pain of others. *Cortex*, 45, 1072–1077.
- Avenanti, A., & Aglioti, S. M. (2006). The sensorimotor side of empathy. In M. Mancina (Ed.), *Psychoanalysis and neuroscience* (pp. 235–256). Milan: Springer-Verlag Italia.
- Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition: Marrying theory of mind and interactionism to study the interplay of Type 1 and Type 2 processes. *Frontiers in Human Neuroscience*, 274, 1–15. <http://dx.doi.org/10.3389/fnhum.2012.00274>.
- Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience*, 3, 1077–1078.
- Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic area. *Proceedings of the National Academy of Sciences United States of America*, 100, 5497–5502.
- Chartrand, T. L., & Bargh, J. A. (1999). The Camaleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893–910.
- Chiavarino, C., Apperly, I. A., & Humphreys, G. W. (2012). Understanding intentions: Distinct processes for mirroring, representing, and conceptualizing. *Psychological Science*, 21, 284–289.
- Cole, J. (2001). Empathy needs a face. *Journal of Consciousness Studies*, 8, 51–68.
- Collingwood, R. G. (1949). *The idea of history*. Oxford, UK: Clarendon Press (revised edition: 1993).
- Corradini, A. (2011). Mirror neurons and empathy: A neuroscientific foundation for morality? In C. Lumer & U. Meyer (Eds.), *Geist und Moral. Analytische Reflexionen für Wolfgang Ienzen* (pp. 117–132). Paderborn: Mentis.
- Dapretto, M., Davies, M. S., Pfeiffer, J. H., Scott, A. A., Sigman, M., et al (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, 9, 28–30.
- Debes, R. (2009). Which empathy? Limitations in the mirrored “understanding” of emotions. *Synthese*, 175, 219–239.
- Decety, J., & Chaminade, T. (2003). Neural correlates of feeling sympathy. *Neuropsychologia*, 41, 127–138.
- De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10, 435–441.
- Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, 11, 86–89.
- Dray, W. (1957). *Laws and explanation in history*. Oxford, UK: Oxford University Press.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, 49, 709–724.
- Gallese, V. (2001). The ‘shared manifold’ hypothesis. From mirror neurons to empathy. *Journal of Consciousness Studies*, 8, 33–50.
- Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 358, 517–528.
- Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror neuron forum. *Perspectives on Psychological Science*, 6, 369–407.
- Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and somatotopic auditory mirror system in humans. *Current Biology*, 16, 1824–1829.
- Heal, J. (2003). Co-cognition and off-line simulation: Two ways of understanding the simulation approach. In J. Heal (Ed.), *Mind, reason and imagination. Selected essays in philosophy of mind and language* (pp. 91–114). Cambridge, MA: Cambridge University Press.
- Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: The Free Press (A Division of the MacMillan Company).
- Hogarth, R. M. (2001). *Educating intuition*. Chicago: University of Chicago Press.
- Hutchison, W., Davis, K., Lozano, A., Tasker, R., & Dostrovsky, J. (1999). Pain related neurons in the human cingulate cortex. *Nature Neuroscience*, 2, 403–405.
- Iacoboni, M. (2008). *Mirroring people. The new science of how we connect with others*. New York: Farrar, Strauss & Giroux.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653–670.
- Iacoboni, M., & Lenzi, G. L. (2002). Mirror neurons, the insula, and empathy. *Behavioral and Brain Sciences*, 25, 39–40.
- Kahneman, D. (2003). A perspective on judgment and choice. Mapping bounded rationality. *American Psychologist*, 58, 697–720.
- Kaplan, J. T., & Iacoboni, M. (2006). Getting a grip on other minds: Mirror neurons, intention understanding, and cognitive empathy. *Social Neuroscience*, 1, 175–183.
- Kim, J. (1984). Self-understanding and rationalizing explanations. *Philosophia Naturalis*, 21, 309–320.
- Kim, J. (1998). Reasons and the first person. In J. Bransen & S. Cuypers (Eds.), *Human action, deliberation, and causation* (pp. 67–87). Dordrecht: Kluwer Publishers.
- Krolak-Salmon, P. et al (2003). An attention modulated response to disgust in human ventral anterior insula. *Annals of Neurology*, 53, 446–453.
- Lawrence, E. J., Shaw, P., Giampietro, V. P., Surguladze, S., Brammer, M. J., & David, A. S. (2006). The role of ‘shared representations’ in social perception and empathy: A fMRI study. *NeuroImage*, 29, 1173–1184.
- Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). Functional imaging of face and hand imitation: Toward a motor theory of empathy. *NeuroImage*, 21, 601–607.
- Minio-Paluello, I., Baron-Cohen, S., Avenanti, A., Walsh, V., & Aglioti, S. M. (2009). Absence of embodied empathy during pain observation in Asperger syndrome. *Biological Psychiatry*, 65, 55–62.
- Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, an emotion. *Personality and Social Psychology Review*, 9, 184–211.
- Pfeiffer, J. H., Iacoboni, M., Mazziotta, J. C., & Dapretto, M. (2008). Mirroring other's emotions relates to empathy and interpersonal competence in children. *NeuroImage*, 39, 2076–2085.
- Pineda, J. A. (Ed.). (2009). *Mirror neuron systems: The role of mirroring processes in social cognition*. New York: Humana.
- Phillips, M. L. et al (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, 389, 495–498.
- Popper, K. R. (1972). *Objective knowledge. An evolutionary approach*. Oxford, UK: Clarendon Press.

- Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1–72.
- Pretz, J. E. (2008). Intuition versus analysis: Strategy and experience in complex everyday problem solving. *Memory and Cognition*, 36, 554–566.
- Rizzolatti, G., & Sinigaglia, C. (2006). *So quello che fai [I know what you are doing]*. Milan: Raffaello Cortina Editore.
- Roganti, D., & Ricci Bitti, P. E. (2012). Empatia ed emozioni: Alcune riflessioni sui neuroni specchio [Empathy and emotions: Some reflections about mirror neurons]. *Giornale Italiano di Psicologia*, 39, 565–589.
- Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, 19, 1354–1372.
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2008). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132, 617–627.
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not the sensory components of pain. *Science*, 303, 1157–1162.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.
- Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 397–420). New York: Cambridge University Press.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences*, 23, 645–726.
- Stüber, K. (2003). Intentional explanation, psychological laws, and the irreducibility of the first person perspective. In A. Bächli & K. Petrus (Eds.), *Monism, Frankfurt a.M* (pp. 255–278). London: Ontos Verlag.
- Stüber, K. (2006). *Rediscovering empathy: Agency, folk psychology and the human sciences*. Boston, MA: MIT Press.
- Stüber, K. (2008). *Empathy*. *Stanford Encyclopedia of Philosophy*, March 31, entered on June 1st 2012.
- Valeriani, M., Betti, V., Le Pera, D., De Armas, L., Miliucci, R., Restuccia, D., et al (2008). Seeing the pain of others while being in pain: A laser-evoked potentials study. *NeuroImage*, 40, 1419–1428.
- Von Wright, G. H. (1971). *Explanation and understanding*. Ithaca, New York: Cornell University Press.
- Waytz, A., & Mitchell, J. P. (2011). Two mechanisms for simulating other minds: Dissociations between mirroring and self-projection. *Psychological Science*, 20, 197–200.
- Wicker, B., Keysers, C., Plailly, J., Rovet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, 40, 655–664.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 358, 593–602.