The many faces of empathy, between phenomenology and neuroscience

Massimiliano Aragona, Georgios D. Kotzalidis, Antonella Puzella

Summary
The definition of empathy differs among the domains which deal with it. Introduced in medicine and psychology in the late 19th-early 20th century, it received contrasting definitions from philosophers and psychopathologists. The neuroscience paradigm of empathy for pain allowed us to identify two components of empathy, one automatic, bottom-up, and one cognitive, top-down. The role of mirror neurons in this context appears to be central. Empathy is influenced by perception of other, closeness, belonging to a social group, and gender, with women empathizing more than men. The areas involved are the self-other distinction areas (dorsomedial prefrontal cortex and temporoparietal junction), the anterior insula, and the anterior cingulate. The activations identified in the brain allow for better understanding the phenomenon, but not to draw a consensus definition. Rather than providing responses, the neurosciences send back to philosophy new, formidable questions to be asked.

INTRODUCTION
The discovery in the mid-nineties of mirror neurons (for an overview see Rizzolatti and Sinigaglia [1]) fostered an upsurge in the interest for the curious phenomenon of putting oneself in another’s shoes, i.e., empathy. This phenomenon, which had always been present in human culture, although in different wording terms, entered the consciousness of Western civilisation with the German term Einfühlung. It was first Johann Gottfried von Herder (1744-1803) who used the term “sich einfühlen”, which literally means feeling one’s way into, to mark identification with an admired object [2]. Later, Robert Vischer (1847-1933), paying tribute to his father Friedrich Theodor Vischer (1807-1887), who had used einfühlen to underline the idealistic view of architectural forms, coined the term “Einfühlung”. He used Einfühlung in the field of Aesthetics to account for the feeling generated when one is exposed to artwork [3]. Some decades later, the same term was used in psychology by Theodor Lipps (1851-1914), who viewed it as an unconscious, instinctual and internally imitative process of fusion between observer and observed object [4]. Lipps extended the concept from a biological organism-object interaction to an interaction between two biological organisms. In 1909, Edward Bradford Titchener (1867-1927) translated the psychological concept of Einfühlung with empathy [5], a pseudo-Greek neologism meaning passion inside, which has never been used by Greeks this way [6], but nevertheless entered our vocabulary internationally.

Mirror neurons are activated not only when an animal (first shown in monkeys, but subsequently also in man) acts, but also when it observes...
the same intentional act by a conspecific [7–10], an act whose intention is understood [11]. Mirror function is not confined to motor acts, but rather constitutes a widespread representational modality of brain function; it extends to several types of response to the environment, including emotional responses and involving their related areas [12–15]. The overlap of brain regional activity during action, emotion or ideation between two persons, one of whom is just observing or imitating the other, offered to clinicians and philosophers dealing with empathy a “brain signature” of the phenomenon and boosted an interdisciplinary debate involving philosophy, psychopathology, and the neurosciences [16].

Despite phenomenologists endorse Husserl’s warning against confounding the simplistic scientific Weltanschauung with the phenomenological method, some attempted to fill the gap by promoting neurophenomenological research programmes. This resulted in unwarranted claims like the discovery of mirror neurons “substantially validate Husserl’s understanding of empathy and the role it plays in embodied intersubjectivity” [17]. Analytic philosophers, together with cognitive scientists and neuroscientists, considered mirror neuronal activity in the empathic process as the proof that the recognition of the other’s perspective is not based on theoretical inferences (the so-called Theory-Theory model) but rather on automatic and implicit activation, in a person, of the corresponding process occurring in the other [1, 18, 19]. According to this view (the so-called Simulation Theory), one’s ability to put oneself in another’s shoes would not be mediated by an elaborate process such as creating a theory about the other’s mind, based on inferences derived from reflecting on what would happen in one’s own self if put in the other’s place. Rather, it would be an innate, embodied simulation of the other’s same action/emotion, spontaneously generated upon observation of a behaviour or facial expression of someone similar to ourselves (like a conspecific), and with the same or similar biological underpinnings [1].

Apparently, the above claims belong to different disciplines, sharing however the following assumption: empathy refers to the same entity, which can be enucleated by means of careful phenomenal description. Once enucleated, the phenomenon has to be explained by its underlying mechanism, and this may be through mirror neuron prepotent activation. This fits Hempel and Oppenheim’s [20] neopositivist epistemological model, which divides explanation in two major constituents, the explanandum, i.e., the sentence describing the phenomenon to be explained, and the explanans, i.e., the class of sentences advocated to account for the phenomenon, which are partly law-like statements and partly sentences stating specific antecedent conditions. In line with the general neopositivist stance that considers the description of the phenomenon to be explained a simple matter of careful observation of what is already given in the objective world, this model analyses the logical and empirical requirements of the explanatory sentences in great detail, while it takes the definition of the explanandum as granted. However, there are no universally accepted descriptions and definitions of empathy, if empathy experts make statements as “There are probably nearly as many definitions of empathy as people working on the topic” [21]. We believe that empathy is still a concept to clarify, worth of being investigated through its neural activity correlates, but not limited by them.

In the first part of the paper we will analyse the lack of agreement in conceptual research on how empathy should be defined and on what it is. We will use a few examples taken from phenomenological philosophy and psychopathology, to show that these differences are due to pragmatic reasons.

In the second part we will show that the neurosciences are revising some early assumptions about brain function during empathy, progressively integrating the more simple implicit, automatic, bottom-up model with a more complex one, involving top-down influences and a temporal differentiation of possible subcomponents of the empathic phenomenon.

Pragmatic reasons for the different definitions of empathy in Phenomenology

The father of phenomenology, Edmund Husserl, developed his views on empathy in his Freiburg in Breisgau years, supported in his work by his assistant Edith Stein. In this phase
of their work they were focusing on the description of the phenomenon of empathy in pure generality, trying to grasp its essence (eidosis). The result of this eidetic study is that the phenomenon of empathy is characterized as an intentional act similar to a perception but sui generis. In empathy as well as in outer perception “the object itself is present here and now” [22, p. 6], but they differ because “I have no outer perception of the [empathized] pain. […] The pain is not a thing and is not given to me as a thing, even when I am aware of it “in” the pained countenance. I perceive this countenance outwardly and the pain is given “at one” with it. There is a close, yet very loose, parallel between empathic acts and the averted sides of what is seen, because in progressive perception I can always bring new sides of the thing to primordial givenness. Each side can, in principle, assume this primordial givenness I select. I can consider the expression of pain, more accurately, the change of face I empathically grasp as an expression of pain, from as many sides as I desire. Yet, in principle, I can never get an “orientation” where the pain itself is primordially given” [22, pp. 6-7]. The conclusion is that in this sui generis intentional act we should distinguish between the fact that we are now empathizing (which we experience as a first person, primordial act) and the empathic content (e.g., the pain that the other is experiencing) which we experience as non-primordial (not as our own but as the other’s experience).

This early account is partly but significantly modified in Husserl’s mature discussion of empathy in the Cartesian Meditations account [23]. Here Husserl characterizes empathy on the basis of a complex line of reasoning that can be summarized as follows: First, by means of the epoche the I (the transcendental “I” of Husserl’s phenomenology, here referred as “the Ego” to simplify) concentrates on its own transcendental sphere of peculiar ownness, and in doing so he finds its own body which is not only an objective body (Körper) but also a living one (Lieb), with kinaesthetic experience and autonomous movements. Second, the other being is perceived in the outer space. However, what is perceived is not “the other Ego [itself], nor [its] subjective processes or [its] appearances themselves, nor anything else belonging to [its] own essence” [23, p.109], because these contents are not giv-
differences, at the point that Waltraut Stein considered the Cartesian Meditations in contrast with Husserl’s first volume of Ideas and with Stein’s On Empathy [22, pp. xiii-xv]. Our point of view is that they are not in contrast but rather the result of different emphasis on different aspects of the concept due to the philosophical questions that had to be answered. In the early account, the focus is on the phenomenological eidetic description of the act of empathy, while in the later one the basic question is focused on intersubjectivity: If I can have direct access only to my intentional states (within my consciousness field), how is it possible to avoid the consequent solipsism? In the early account it is assumed that “the phenomenon of foreign psychic life is indubitably there” [22, p. 5], while in the later one it is this evidence itself that is in need of justification. This difference in the questions is in our view responsible for the differences in the resulting phenomenological description of empathy.

Karl Jaspers imported a phenomenological approach into the realm of psychopathology. In his General Psychopathology [24], empathy refers to the act of understanding (Verstehen), defined as the act of grasping psychic events “from within”. “Since we never can perceive the psychic experiences of others in any direct fashion” – Jaspers suggests – “There has to be an act of empathy, of understanding”. Accordingly, “The first step, then, is to make some representation of what is really happening in our patients, what they are actually going through, how it strikes them, how they feel”. This act of intuitive apprehension of the other’s lived experiences (Erlebnisse) which characterises Jaspers’ phenomenology is quite different from Husserl’s phenomenological empathy. We think that the main reason for this difference is that Jaspers’ question was radically different from Husserl’s. Jaspers needed to render subjective experience scientific, against all brain mythology [25]. To do so, empathic understanding had to take into account all the complex material regarding the other against the background of one’s own representations. Hence, Jaspers’ basic question was: How can I scientifically know the subjective experience of my patients? In so doing, Jaspers adopted a clinical, pragmatic viewpoint that detached itself from Husserl’s purely philosophical standpoint.

Empathy and the Neurosciences

Meanwhile, the neurosciences were adopting their own scientific methods to investigate cognitive phenomena. In early years, investigations would limit themselves to measuring galvanic skin response or non-specific brain activity, such as the one that may be deduced through the electroencephalogram or stimulus-evoked potentials, and they would combine this with the use of questionnaires, that allow for factorial analysis, thus scientific analysis. The advent of functional magnetic resonance (fMRI) and the discovery of the mirror neurons enabled us for the first time to “see” the brain in action and a representational network to reveal itself. This stimulated philosophical questions among the neuroscientists and it is likely to form bridge between scientific and philosophical constructs.

The mirror neuron concept was originally limited to imitation (thus, intention-informed) of motor action; mirror neurones were first identified in areas which were adjacent to the groups of neurons that controlled the action being simulated or observed [10, 11]. A study investigating mirror neuron function in empathy shifted the focus from simple motion to emotion [12], showing that mirror functioning is a widespread modality of brain functioning extending to several areas.

According to the Parma group’s Simulation Theory, being in a relationship with another person is not mediated by a theory of the other’s mind, but rather by an implicit, inbuilt, automatic process which goes without logical-deductive procedures, although they do not exclude the possibility that it is modulated by such procedures; this would involve that it is process-driven rather than theory-driven [1, 18]. Thus, putting ourselves in another shoes would activate the corresponding process occurring in the other while performing the same activity or experiencing the same emotion.

Due to the expansion of the empathy concept and the lack of a consensus definition, we will here limit ourselves to the consideration of one single paradigm, i.e., functional neuroimaging
in empathy for pain. Pain constitutes a universal experience in the animal kingdom, with few inter-cultural differences and a known representational brain matrix that may be easily shared, comprising parts of the thalamus, insula, cerebellum, posterior and rostral anterior cingulate cortex, and their connections, originally called neuromatrix; these areas are activated by pain, although they are not specific for pain.

Singer et al. [14] used fMRI to assess activated brain areas during actual painful stimulus perception in women versus areas activated when the painful stimulus was delivered to a significant other (partner/spouse). Bilateral anterior insula, rostral anterior cingulate cortex, extrastriate visual cortices, brainstem, and cerebellum were activated in both self and other pain, while posterior insula, caudal anterior cingulate area, and contralateral secondary somatosensory and sensorimotor cortices were more activated in self pain. Anterior cingulate and insular activations correlated strongly with rating scale-assessed empathic tendency.

To investigate gender differences, Singer et al. [15] exposed men and women undergoing fMRI to the view of a fair and an unfair player engaged in an economic game and then subjected both participants and players to painful stimulation. Both men and women activated empathic brain areas when watching fair players being subjected to pain, but the view of an unfair player receiving pain split the response, with women activated the same areas, albeit less, and men showing activation of the basal ganglia and nucleus accumbens, which are pleasure/reward areas. Empathic tendencies again correlated with fronto-insular and anterior cingulate activation in both men and women.

The importance of perceived intentionality was underlined by a series of studies of Aglioti’s Roman group. They used transcranial magnetic stimulation (TMS) to show pinched muscle-specific reduced evoked potential amplitude correlating with sensory, but not emotional ratings of pain, in the sensorimotor cortex, which varied significantly when the participant sees a hand being pinched compared to when it is touched with a cotton swab or when a tomato is being pinched [26]. Sensorimotor responses were not affected by the set of instructions given to the participant in a subsequent study [27]. According to these authors, “simpler” empathy is related to the mapping of an external sensory stimulus, and “complex” empathy to affective tuning, which is keen to emotional sharing. Another TMS/motor-evoked potential study showed larger effects with left motor cortex stimulation [28]. Another study using somatosensory evoked potentials and simultaneous electroencephalographic (EEG) recording showed that the primary somatosensory cortex (S1) is involved in both self-pain and in other-pain and touch mapping, i.e., in extrapolating somatic features in the context of social interactions [29]. Subsequently, the same group used the same own-hand/other’s hand/tomato paradigm to rate pain matrix response to own and other’s pain, delivering the same type of pain to both participant and stranger model [30]. This study showed that laser evoked potentials were similar for other’s pain, but the extent of the pain matrix was modulated by pain intensity rating, i.e., empathy for another is modulated by own experience, thus underlying the importance of first-hand experience.

A similar TMS study showed somatomotor mirror responses to other’s pain to be modulated by observant’s empathic attitude [31]. Trait empathy correlated with the extent of mirror neuron activation in the observant, meaning that the reduced corticospinal excitability accompanying the vision of pain inflicted to another person, which is used by people to reduce their pain perception specifically to the muscle being pinched, is activated in social contexts as a possibly learned, but currently embodied phenomenon. Another study showed that the “freezing reaction” aimed at soothing pain in the painfully-stimulated muscle, is paralleled by an “escape reaction” in the contralateral muscle, i.e., increased corticospinal excitability when the pain is inflicted to the side opposite to that from which motor evoked potentials are recorded, and this was interpreted as an automatic response [32].

Backing Singer’s et al. [14, 15] original observations, other fMRI studies, using photographs of painful stimulation rather than actual pain, confirmed that anterior insula and anterior medial cingulate activation mediated empathic response to pain [33–35], while psychopathic traits in youths, which are generally related to lack of...
empathy, decreased activation in rostral anterior cingulate and other areas [36]. Activation correlated with both estimated pain intensity and empathy. However, actual pain and photographs may have different valences; furthermore, when cartoons are used rather than photographs, besides anterior cingulate and paracingulate cortex, and right middle frontal gyrus activations, left parietal cortex, postcentral gyrus, and occipito-temporal cortex activations were additionally displayed, pointing to a more rationalized response and a lower empathy for pain, a fact pointing to a top-down model of empathy for pain [37].

Among social factors of empathy for pain, a racial bias was shown involving increased insula, anterior frontal cortex, and anterior cingulated cortex activations in the same-group, compared to the other-group condition [38, 39]. However, social integration tends to reduce this bias [40]. Another study showed stronger reduction of muscle-specific corticospinal excitability with same-group [41].

Neuroimaging studies of social perception and empathy showed simultaneous processing of self-other distinction with increased inferior temporal cortex activation [42]. Similarly, investigating the influence of cognitive appraisal in empathy for pain, the automatic response was modulated by self/other distinction and valence attribution areas [43, 44]. Activations in the self- and other-perspective were similar, but the former showed stronger activation in the left parietal cortex, amygdala, insula, and anterior medial cingulate area, while the latter in the right parietal cortex. It was apparent that the self-perceived involvement emotional response to threat. These studies showed empathy not to be a purely automatic process, based only on somatic/motor tuning and resonance, but also to be influenced by top-down mechanisms independently generated in the cortex after stimulus evaluation. Top-down regulation of empathy was identified in another study [45]. Besides activation of the anterior insula and the anterior medial cingulate cortex, empathizing with the pain of a “different” other recruited areas involved in self-other distinction (dorsomedial prefrontal cortex) and in executive functions (dorsolateral and right inferior frontal cortex). Furthermore, the effective connectivity between the insula and these frontal areas was enhanced by the empathizing process, thus confirming that neural structures activated during empathy for pain are similar for making inferences about the affective state of “same” and “different” others.

Fan and Han [46] proposed an empathy model consisting of an early, bottom-up component, which is linked to emotional sharing and self-distress for viewing a conspecific suffering, and another late, top-down, which relates to cognitive evaluation. A further study examined the effects of perspective-taking [47], showing that the early component of the pain judgement task was left unaffected, while the late component was much reduced by the other-perspective. Another fMRI perspective-taking study [48] investigated the influence of empathy by relatedness. Participants were asked to rate the pain imagined to be inflicted upon their own limbs, to those a loved one, and to a stranger. The bottom-up, automatic component of empathy activated the anterior insula and anterior cingulate, as expected, but was significantly stronger for self and loved ones, compared to strangers. Stranger-pain activated the self/other distinction areas, namely, the superior frontal gyrus and the temporoparietal junction. Pain to a loved one deactivated the temporoparietal junction and this correlated with the closeness of the participant to the loved one, while in the stranger condition there was a negative effective connectivity between insula and temporoparietal junction and a positive effective connectivity between temporoparietal junction and superior frontal gyrus. This study showed that relatedness and affective involvement with a person reduced the self/other distinction during the bottom-up processing of empathy, while the self/other distinction predominated in empathy for a stranger, whose pain was likely to receive low ratings.

Despite fMRI studies are not to be taken as the proof, but rather than a hypothesis-setting start, it should be underlined that they are usually of insufficient sample size for conclusions to be drawn; meta-analyses are needed to confirm whether the apparent consistency of these studies is real. In fact, a recent meta-analysis identified both commonalities and differences in brain activities between directly experienced pain and empathy for other’s pain and related them to different methodologies [49]. However, when com-
combined together, neuroimaging studies, despite differences in methodology, agree on the existence of gender differences (which are in turn probably influenced by personality differences, as shown in the case of prosocial predispositions [50]), racial bias, perspective-taking (self-other distinction), and relatedness during empathizing. In one word, it relates to self-representation and identification. It is likely to represent an embodied, automatic phenomenon which is modulated by experience, including evolution, and possess a top-down, cognitive arm which ensues in forming social relationships and shapes the formation of cultural groups and social values. This way it poses formidable philosophical questions, to feed-back to philosophy and clinical psychopathology, to ensure a continuous dialogue. In fact, Karsten R. Stueber took the baton from simulation theorists and theory theorists to emphasise the current limitations of science and the paucity of philosophical debate and to propose new strategies for distinguishing between basic and re-enactive empathy [51].

CONCLUSIONS

There is not a single definition of empathy. The explanandum being heterogeneous, we cannot expect to discover a common mechanism explaining it. What is meant by empathy depends on the basic scientific/clinical/philosophical question. Within the neurosciences, empathy is shifting from a simple automatic bottom-up process to a more complex phenomenon in which different mechanisms (either bottom-up and top-down processing) are involved.

REFERENCES