FEDERICA MADONNA
Università degli Studi di Cassino e del Lazio Meridionale
f.madonna@unicas.it

THE OVERLAP OF MINDREADING AND METACOGNITION: AN ELIMINATIVIST EXPLANATION OF COGNITIVE EMPATHY[1]

abstract

This paper suggests that the concept of cognitive empathy, as currently presented in much of the literature, may rest on conceptually fragile assumptions. The argument is developed in three steps: a reconstruction of the historical misunderstandings behind the idea of empathy; a comparison between mindreading and metacognition; and a proposal for an alternative view of intersubjectivity based on what is defined here as a "relational void". Rather than demonstrating definitive claims, this contribution aims to open a critical reflection on the theoretical status of cognitive empathy and its possible redundancy when viewed through the lens of metacognitive processes.

keywords

consciousness, cognitive empathy, mindreading, metacognition, relational void

1 I thank Prof. Andrea Lavazza for the valuable suggestions he provided me during the writing of this article.

Phenomenology and Mind, n. 29 - 2025
DOI: 10.17454/pam-2902
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Introduction

In recent decades there has been much talk about *empathy*; at times it has been placed in a neurobiological context; others in a relational context; others yet in a cognitive context. It is no coincidence that attempting to provide a *clear and distinct* definition of the phenomenon is fraught with difficulty, as there are as many definitions as there are scholars engaged in the debate[1].

Indeed, the term "empathy" has undergone a remarkable semantic evolution, resulting in a series of frequently divergent definitions. Some authors have given priority to its affective aspect[2], describing it as the ability to resound with the other on an emotional level. Others have emphasized its cognitive dimension, understood as the ability to take another's perspective or infer mental states[3]. In addition, more recent contributions have sought to integrate neurobiological, social and moral components, emphasizing the adaptive function of empathy[4]. While this theoretical plurality enriches the debate, it also runs the risk of creating conceptual confusion, especially if a definition is assumed to be universally agreed upon. It is precisely this complexity that calls for a critical re-examination of the concept of cognitive empathy as it is commonly understood.

What holds this conceptual tangle together is, arguably, the relationship between two individuals (henceforth: an I relating to a You) which is capable, through this strange phenomenon, of "coming into contact". The "strangeness" could be explained in a very haphazard mixture of aspects that, in the very history of the concept, have become intertwined without any longer providing conceptual clarity as to what these very aspects were supposed to explain. We refer to the fact that the very sketchy explanation of empathy as "the ability of the I to experience first-hand the emotions of the You," the result of which

1 Cf. J. Decety, W. Ickes (a cura di), *These Things Called Empathy: Eight Related but Distinct Phenomena,* from *The Social Neuroscience of Empathy*, MIT Press, Cambridge, 2011,

2 Cff. N. Eisenberg, *Empathy and Sympathy*, in Michael Lewis, Jeannette M. Haviland-Jones, and Lisa Feldman Barrett (eds), *Handbook of Emotions*, Guilford Press, New York, 2008, pp. 677-691.
E. Aaltola, *Affective empathy as core moral agency: psychopathy, autism and reason revisited*, in "Philosophical Explorations", 17(1), 2013, pp. 76-92.

3 Cff. S. Baron-Cohen, Zero Degrees of Empathy: A new theory of human cruelty and kindness, Penguin, London, 2025; J. Decety and W. Ickes (eds), *The Social Neuroscience of Empathy*, MIT Press, Cambridge (MA): 2009.

4 Cff. B. Cuff, S.J. Brown, L. Taylor, D. Howat, Empathy: a review of the concept, in "Emotion Review" Vol. 8, Issue 2, 2016, pp. 144-153; C.D. Cameron, P. Conway, J.A. Scheffer, Empathy regulation, prosociality, and moral judgment, Curr Opin Psychol., Vol. 44, 2022, pp. 188-195.

would be that the I and the You feel the same suffering or joy conveyed by the interlocutor at that moment, has been enriched, in a symbolic way from the 1990s onward, with neuroscientific aspects (in this regard, mirror neurons represent the most relevant aspect) to explain this particular inter-relationship; the consequence was that from *empathy* we have ended up talking about *mentalization* or the ability to "read [neurobiologically, Ed.] the mind of the You"[5]. This distinguishes it from empathy, which is understood as an emotion felt only by the I, regardless of the situation, the context, and whether the interlocutor is present or not[6].

This "evolution" has occurred in stages so much so that, following Martin Hoffman's[7] timely analysis, in the current debate the "empathy problem" seems to have found its own universally shared subdivision: on the one hand, basic empathy or primary aspect; on the other, cognitive empathy or secondary aspect and its possible relationship to how morals came to exist. In the first case, referring to the relentless development of cognitive neuroscience in recent decades, it is understood as the set of all the neurobiological mechanisms underlying the phenomenon (experiencing first-hand the emotions of the other on oneself"); in the second, emphasis is placed on a different and founding aspect of the phenomenon, which involves the cognitive effort of the I to understand the experience which the You communicates to it, which in itself has the effect of experiencing first-hand the emotions of the You; this is referred to as secondary or cognitive empathy. An outgrowth of this aspect is the interest that has arisen around the idea that there may be a relationship between the role of empathy and the emergence of morals[8], extending the question of whether altruism can identify a pre-social form of empathy, whether they are two distinct phenomena, or whether one is the cause of the other.

While basic empathy, therefore, has initiated a field of research aimed at understanding which areas of the brain are involved in giving rise to empathy itself, cognitive empathy is the subject of numerous interpretations, which focus on how this cognitive effort of the I can take place in order to understand the other, the result of which is a cognitive form capable of "entering" into the interlocutor's mind and innermost experience. The characteristics of cognitive empathy can be summarized in three: the processing time of the message; the cognitive effort to decode it; and, lastly, the psychic distance between the two interlocutors[9].

As can be seen, in the cognitive dimension, the interpretation of the empathic phenomenon is completely separated from the neuroscientific aspect and, while up to the early 2000s this "field of research" focused on the different facets of the phenomenon itself, identifying its different characteristics and nuances, in the last twenty years (thanks to the continuous development of neuroscience) the field of research is not "clear", uniting both the primary and secondary aspects. This fusion has given rise to a real phenomenon in itself, *mentalization*, which allows one of the two interlocutors to "enter into the mind of the other, reading its intentions, desires and beliefs"[10] through specific neurobiological mechanisms for mindreading"[11], and then recreate them-neurobiologically and cognitively-into oneself,

5 Cf. R. Almeida Et Alii, *Neural Correlates of empathy: a systematic review of event-related potentials studies in perceptual tasks,* in "Brain. Sci.", 14(5), 504, 2024

6 Cf. E. Aaltola, *Affective empathy as core moral agency: psychopathy, autism and reason revisited*, Philosophical Explorations, 17:1, 2014, pp. 76-92.

7 Cf. M. Hoffman, *Empathy and Moral Development. Implications for Caring and Justice,* Cambridge University Press, Cambridge, 2010.

8 Cf. C.D. Cameron, P. Conway, J.A. Scheffer, *Empathy regulation, prosociality and moral judgment*, op.cit.

9 M. Hoffman*, Empathy and Moral Development*., op. cit.

10 Cf. J. Stietz J., E. Jauk, S. Krach, P. Kanske, *Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research*. Front. Psychiatry, 10:126, 2019.

11 D. D. Hutto Et Alii, *Editorial: Social Cognition: Mindreading and Alternatives*, in "Rev. Phil. Psych.", 2, 2011.

opening up perspectives on how to "project towards oneself" a thought which originated in the other's mind and "transport" it into one's own (the problem of metacognition).

Thus, mindreading represents the most recent evolution of the "empathy phenomenon" to date, suggesting that the multiple successive interpretations of the "I-You" relationship suggest that the concept of "cognitive empathy" itself is structurally fragile, opening up the possibility of its critical revision.

In fact, the intention of this article is precisely to suggest that a way forward in this direction is possible by critically exploring alternatives that would lead one to think that some interpretive models used to explain the phenomenon (e.g., those that have relied on the use of folk psychology or the use of the concept of mind) may have contributed to the construction of illusory representations about the role and explanation of what cognitive empathy[12] is; this could lead to the delineation of different models of the mental that, in a likely explanation, lead to the introduction of the concept of "relational void".

In order to achieve this goal, a number of key steps will be necessary. In particular, the first section will explain the emergence and revival of the concept of empathy up to its declination as "cognitive empathy", and the second section will focus on the transformation of the concept of cognitive empathy into "mind reading", clarifying the reasons why a discussion has started in the contemporary debate on how it is possible to "think about one's thoughts (metacognition).

This is done in order to provide the conceptual premises to address the ultimate purpose of this paper, which is to suggest that it is plausible to initiate a critical revision of the concept of cognitive empathy itself.

It is important to clarify that this paper does not intend to question every possible meaning or declination of cognitive empathy. The analysis focuses specifically on certain theoretical formulations which currently prevail, such as that of Martin Hoffman, which we believe contain elements of ambiguity or conceptual redundancy. We do not rule out the possibility that other approaches – especially phenomenological and embodied ones – may offer more coherent readings of the phenomenon, but they are not the specific focus of this discussion[13].

Thus, in the next section, we will hypothesize that mentalization is a conceptual duplication of the phenomenon of metacognition theorized by John Flavell in the 1980s, since it overlaps perfectly with "mind reading," in order to draw the reader's attention in the fourth section to a naturalistic reading of the mind, anticipating what is meant by "relational void," i.e., an alternative explanation of intersubjectivity that will be made explicit only in the conclusion.

2. The origin of the misunderstanding: from *sympathy* to mindreading

The origin of the concept of cognitive empathy, as it has been defined, is relatively recent, since it has been included in scientific debate only since the 1950s, following a revival of the concept by philosophers Willard van Orman Quine, Hilary Putnam and Robert Nozick. Subsequently, it underwent considerable changes until its final transformation into "mindreading".

Understanding these changes, starting from their historical origin, involves having a clear idea of the common thread that has formed the trail on which various misunderstandings and various answers have been provided in order to address old dilemmas, attempting to

12 Cf. B. Cuff, S.J. Brown, L. Taylor, D. Howat, *Empathy: a review of the concept*, op. cit.

13 The reference to Hoffman is understood here as representative of a widespread theoretical line, particularly in the fields of developmental psychology and cognitive neuroscience. It is obvious that the position advocated in this paper does not claim to apply to every notion of cognitive empathy, but merely specifically criticizes this theoretical approach and its developments with respect to mindreading and meta-representation.

completely abandon any recourse to metaphysics. Such claims are reflected in the different steps in the evolution of the phenomenon, beginning precisely with David Hume, who, both in his *Treatise of Human Nature*[14] and in his *Enquiry Concerning the Principles of Morals*[15], explained that the foundation of morals consisted of *sympathy* first, which is then transformed into *natural benevolence*.

In the first work, the Scottish philosopher claimed that *sympathy*, in addition to defining the mutual action of things which affect one another, is able to make us *enter so deep into the opinions and affections of others whenever we discover them*[16] precisely because it represents the internal principle of the passions by which the transition from *ideas to impressions* is achieved. An epistemological tool that, excluding any "emotional fusion" between interlocutors, originates in itself and results in *imaginative participation in the lives of others*[17], providing a "primary" knowledge of the other. In *Enquiry* these aspects of *sympathy* are extended by Hume himself to a universal explanation as to why in the human there is a feeling of benevolence, of "openness to intersubjectivity", which explained why two or more subjects are able to cooperate, converse, and interact.

It is not a methodological error to start precisely from the Scottish philosopher in order to understand the origin of today's cognitive empathy, since in his philosophy there are several points of junction with modern conceptions of empathy, both in its relation to morals and in the relationship between two or more individuals. The latter aspect, completely secondary for Hume as in the entire current of the eighteenth century, was given special attention only as a result of the interpretation of the meaning of *ein-pathos* ("entering into the passions") deduced from the philosophy of Friedrich Schleiermacher who, by conveying the emotional aspect of sympathy to the field of art, argued that in all works of art the artist transposed a part of himself and his "authentic" I onto the canvas, therefore the viewer "knew" and "emotionally felt" the artist's interiority. Only as a result of such a "change of direction" did the intersubjective aspect enter the collocation of the concept of empathy which, transformed completely from Hume's real intentions, became, both with Theodor Lipps and Carl Rogers, an explanation of why one understands the other by experiencing first-hand the interlocutor's experience.

Also in this respect there has been a controversial methodological and interpretative approach on the evolutionary steps of the phenomenon, as today's debate recognizes Lipps himself as the "founder of empathy"; upon a careful reading of his works, it is noted that the aim that moved the philosopher from Leipzig was not to provide an explanation of intersubjectivity, but to "give order to the world" as Immanuel Kant himself, through the categories, did in his *Critique*.

Sharing the views of contemporary philosopher Dan Zahavi, Lipps seems to explain empathy as the result of a motor mechanism rather than as an explanation of the mental life of others[18]. Following the criticism of Edith Stein or Edmund Husserl, the German philosopher

14 Cf. D. Hume, *A Treatise of Human Nature, Paperback, Firenze, 2023.* We point out that the literature on the interpretation of *sympathy* in Hume offers two explanations: one argues that the latter can correspond to a kind of "emotional fusion" between different individuals who may not live in the same social context; the other rules out such an interpretation, since in both the *Treatise* and the *Enquiry* Hume precisely distinguishes sympathy/natural benevolence from emotional contagion, pointing out that these are two different aspects. Here we agree with the latter interpretation.

15 Cf. Id., *An Enquiry Concerning the Principles of Morals, Paperback, Firenze, 2023.*

16 D. Hume, *A Treatise of Human Nature*, op. cit., pp. 334-335.

17 I. Cappiello, *La morale della simpatia in David Hume*; Liguori Editore, Napoli, 1983, p. 87.

18 Cf. D. Zahavi, *Empathy and Other-Direct Intentionality*, in "An International Review of Philosophy", Vol. 33, n. 1, 2014, pp. 129-142.

sees a discrepancy between the phenomenon to be explained (empathy) and the phenomenon actually explained, because to resort to the "commonality of feeling" between the I and its interlocutor simply by claiming that it is due to an "emotional resonance" – as Lipps claims – is not to have explained the difference between recognizing another's emotion and knowing it. They are two completely different epistemic aspects, says Dan Zahavi, who argues that since empathy is a phenomenon that can be interpreted in different ways, to the point of being almost a polysemantic term, it would be useful to always take into account in its use the phenomenological difference between knowing others, oneself and external objects, so that we try to keep the different levels of explanation of the empathic phenomenon separate[19], although we may wonder if it would not be more honest to abandon the use of the term altogether[20].

Continuing with the analysis of the concept, a further transformative step of cognitive empathy derives precisely from analytic philosophy, especially from Quine who, through the use of "radical translation," with the precise intention of "overcoming" Carl Gustave Hempel's Deductive Nomological model by criticizing it, recommended giving philosophical thought to the concept of empathy, understood as a real cognitive tool capable of providing knowledge on the desires or volitions of the other. This operation of restoring the concept was endorsed by Putnam and Nozick themselves: the former by providing *simplicity and plausibility*[21] ; the latter, by paving the way, following the publication of the paper by the two ethologists David Premack and Guy Woodruff [22], for what would become "mindreading".

In such illustrative passages, a gradual transformation of the concept of empathy is evident: from *sympathy* to mentalization, suggesting the interpretation that this phenomenon responds to questions left unresolved in modern times, that is when, symbolically from Descartes, the examination of the problem of the nature of the mental, whose reflections are still debated today. An example of this is Gustave Hempel's Nomological-Deductive model, which – by the very nature of the analytical approach – completely excluded the metaphysical and emotional aspects from the understanding of the phenomena since both were unable to respond to the rational formalism of syllogism. In this context, while, on the one hand, Quine had the insight to "break out" of this schematism, restoring the concept of empathy, on the other hand, the latter was useful to Quine himself as a purely natural element to explain the relational aspects of the mental experiment.

In this way, by providing empathy with the nature of an emotion, and placing it in the realm of the pragmatist, Quine was able to respond to Cartesian dualism (are we thought or matter or both?) by not resorting to any speculative metaphysics on the *cogito: res cogitans*, that is, it was naturalized. It is along these interpretative lines that, subsequently, in order to avoid any form of mentalism, cognitive empathy was transformed from an "embodied instrument" of the person into an "objectified instrument" with Wilfrid Sellars' theorization of the "Myth of Jones," first, and of folk psychology, later. A conceptual framework that, by resorting to what is most "existential" and pragmatic, is explained as:

> all the platitudes you can think of regarding the causal relationship of mental states, sensory stimuli and motor responses [...].

19 Id., *Beyond empathy. Phenomenological Approaches to Intersubjectivity*, in "Journal of Consciousness Studies", 8, n. 5-7, 2001, pp. 151-167.

20 *Idem*.

21 Cf. H. Putnam, *Mind, Language and Reality*, in "Philosophical Paper", Vol. 2, Cambridge University Press, Cambridge, 1995.

22 Cf. D. Premack, G. Woodruff, *Does the Chimpanzee Have a Theory of Mind?* In "Behavioral and Brain Sciences", n. 4, 1978, pp. 515-629.

> And also, all the platitudes to the effect that one mental state falls under another [...]. Perhaps there are platitudes of other forms as well. Include only platitudes which are common knowledge among us: everyone knows them, everyone knows that everyone else knows them, and so on (Lewis, 1972, p. 256),

transforming cognitive empathy into a valid stratagem to avoid the question of the nature of the mental.

Despite a controversial conceptual history, empathy has provided the current debate with the incipit of several fields of study, which can be summarized in five major directions:

- *Temperamental or personality variables related to empathy-related responding*, in which attention is paid to the degree of empathic intensity experienced by the agent placed before a victim's distress, whose variables are their temperament and personality;
- *The development of empathy-related responding*; in fact, for the past few years, studies aimed at the study of the emotional-empathy development of minors have arisen from their early stages of cognitive development;
- *The relation of empathy-related reactions to social behavior, including prosocial behavior, aggression and social competence*; taking up and reproducing Milgram's experiment, attention has been paid to how the perpetrator's empathy is annihilated before the authority figure or how, on the other hand, empathy may be the motivation for the agent to engage in prosocial behavior, perhaps at the basis of altruism[23];
- *Gender differences in empathy-related reactions*, in whose research an attempt is made to provide an explanation of why one gender would be more inclined to feel empathy than the other;
- *Socialization correlates*, in which the limitations to which empathy is subjected are highlighted, referring, in particular, to the different response of the agent when faced with a victim's situation of suffering: from a close family member or relative to a complete stranger.

The latter, stigmatized (in a few lines here) by Nancy Eisenberg[24], summarize the heart of the problem[25]: despite the fact that it is almost impossible to define empathy unanimously, an attempt is made to make the entire theoretical construction of the phenomenon plausible by making the explanation of what it is plausibly true.

3. From cognitive empathy to mentalization

It has already been said that the concept of empathy has undergone several shifts in meaning from Hume's original intent to contemporary times, and it has also been mentioned that in today's debate, as a result of the ongoing neuroscientific developments of the last thirty years, it has been agreed to separate the meaning of empathy from a basic to a cognitive interpretation; a "separation" that, if for nearly twenty years it was the pragmatic basis of a certain explanation of cognitive empathy (the "embodied simulation" of folk psychology), in recent years this – unanimously recognized – distinction is constantly "betrayed" in an ongoing attempt to provide simplicity and plausibility to the phenomenon itself. This suggests that the main problem of cognitive empathy is based precisely on the network of cascading processes that it sets in motion to ensure that the steps related to "understanding

23 D. L. Krebs, *Empathy and Altruism*, in "Journal of personality and social psychology", 1975, XXXII, pp. 1134-1146.
24 M. Lewis, J.M. Haviland-Jones, *Handbook of Emotions*, Guildford Press, New York, 2018, pp. 677-691.
25 L. Wispè, *The distinction between sympathy and empathy,* in "Journal of personality and social psychology", Vol. 50, n. 2, 1986, pp. 314-321.

and feeling the emotional states of others" are followed; which would make intersubjectivity possible: "entering into the mind of the other" in order to acquire an "epistemic" knowledge of it; "recreating it in oneself"; experiencing the same emotion of the other's subjectivity cognitively and sensorially.

An example of a "basic" approach to explaining empathy is that of Frederique De Vignemont and Tania Singer, who argue that empathy could be defined as a "modular", unconscious and implicit response to the social situations in which the individual interacts. Sharing the same neural mechanisms, the empathic response is activated with the help of two different modular pathways, each with a specific function: a first mechanism involves a conscious, voluntary use of the empathic response, taking advantage of the subject's conscious control of the affective response to the other's emotion; a second mechanism involves an implicit, unconscious, immediate assessment of the other's emotional situation, capable of influencing the emotional strength of the response. In both situations, according to the two authors, the empathic response is activated directly and automatically, emphasizing that the latter has a purely social and epistemological role: that is, it is able to provide a direct estimate of the possible future actions taken by the interlocutor (social role), giving an explanation (epistemological role). A limitation of such an instrument, also according to the authors, is the sharing of the same experiential and cultural background[26].

What is evident in De Vignemont and Singer's explanation of empathy is certainly the value of the constantly developing neuroscientific research that attempts to identify the brain areas or mechanisms involved in the phenomenon; what does not seem to be the object of analysis, however, is an "ontological" explanation of empathy itself: in fact, the two authors start from the assumption that the phenomenon exists and that it can have certain or certain other explanations.

Alternative approaches to intersubjectivity – such as phenomenological[27] or direct perception approaches[28] – offer different interpretations of the relationship between self and other that are not based on simulation or meta-representational models. This paper, however, takes a different perspective, aiming to explore the conceptual fragility of the notion of cognitive empathy as it is commonly understood. This does not mean that such approaches do not deserve careful consideration in relation to the themes of this article, which can be done in a subsequent in-depth study.

It should be acknowledged, however, that alternative approaches to the theory of mind – particularly those of direct perception – have proposed different models for explaining intersubjectivity, freeing themselves from both simulation and meta-representation. Authors such as Shaun Gallagher[29], Pierre Jacob[30] and Joel Krueger[31] argue that understanding the other can occur through direct, embodied and contextually situated access, thus reducing the role of "internal" and inferential mental processes. Although this paper does not explore these perspectives in depth, they pose a relevant challenge to the automatic identification of empathy with complex cognitive operations and deserve careful consideration in future

26 Cf. F. De Vignemont, T. Singer, *The empathic brain: How, When and why?*, in "Trends and Cognitive Sciences", Vol. 10, n. 10, 2006, pp. 435-441.

27 Cf. D. Zahavi, *Beyond empathy*. op. cit.

28 Cf. S. Gallagher, *Direct perception in the intersubjective context*, in "Consciousness and Cognition", 17, 2008, pp. 535–54

29 Cf. S. Gallagher, *Direction perception in the intersubjective context*, in "Consciousness and Cognition", 17, 2008, pp. 534-554.

30 Cf. P. Jacop, *The Direct-Perception Model of Empathy: A Critique,* in "Review of Philosophy and Psychology", 2 no. 3, 2011, pp. 519-540.

31 Cf. J. Krueger, Direct Social Perception, in A. Newen, L. De Bruin, S. Gallagher (EDS), *The Oxford Handbook of 4E Cognition*, Oxford University Press, Oxford, 2018, pp. 301-320.

work, especially in relation to the critique of mentalization as a conceptual duplicate of metacognition.

Continuing the analysis, the moment when cognitive empathy symbolically became "the ability to enter the mind of the other to interpret their intentions, beliefs and desires" took place, as said above, with the publication of the paper by the two ethologists Premack and Woodruff in 1978, in which the two authors questioned whether chimpanzees also had an ability to understand the intentions of others, consequently adopting a behavior appropriate to the stimulus understood.

On account of the peculiarities of the task implied by this ability, people began to speak about a real *theory of the mind*, which could be applied to nonhuman animals as well. Hence the origin of *mentalization*, i.e., a further transformation of the common meaning of cognitive empathy to indicate, specifically, a particular ability, neurobiologically inherent in humans and not only in the human species, to *actually* "enter" the mind of the other; from a "mere" cognitive effort that allowed the I to "understand" the You, we moved on to the identification of a "functional module" of the entire brain capable of *letting one's mind enter* that of the other.

From a strictly speculative point of view, the originality of the two ethologists' essay was to fit into the progressive construction of the *empathy stratagem* to solve the problem of intersubjectivity; that is, there was a need for an element that was able to explain why an agent, situated within an already given context (*folk psychology*), was able to "understand the other". The use of means which were already known (shared beliefs) was not sufficient, as one sought to provide a rational justification for the kind of epistemic knowledge that empathy could provide; without "scientific" support, the latter simply appeared to be mere intuition. Therefore, Premack and Woodruff's paper provided the "lifeline" from the collapse of the *Folk Psychology* theory and its countless internal facets, which can be summarized in the two great families of *Theory of the Theory* (TT) and *Simulation Theory* (ST).

TT refers to the explanation of intersubjectivity by using a set of "common laws" (*folk*) through which beliefs, desires and volitions are attributed to the other than Self: *folk Psychology*, therefore, is equivalent to an "already given" theoretical framework used by the agent through which he can understand that of the interlocutor[32].

In the case of ST, on the other hand, "common sense psychology" is interpreted as a "false theory", referring to completely non-existent objects (mental states); therefore, the explanation of intersubjectivity is completely entrusted to the brain processes of the agent himself, which are able to initiate "internally" – in a "simulated" manner – the "understanding" of others' mental states thanks to special cognitive mechanisms which, from the 1990s onward, were hypothesized to be mirror neurons[33] along with other areas of the entire encephalon[34].

Although the simulation theory has been strongly advocated by many to explain the way in which we "understand" or interact with each other, the "ST paradigm" has been in crisis for several years due to severe criticisms of the way it works. One of them is the one put forward and shared by Shaun Gallagaher[35], according to which ST is based exclusively on the neural process (high or low level) by which the I, drawing on its own experiential background, "reproduces in itself" the emotional state of the other, experiencing the same feeling.

32 Cf. K.R. Stueber, *Rediscovering Empathy: Agency, Folk Psychology and the Human Sciences,* MIT Press, Cambridge, 2010., 2010.

33 Cf. G. Rizzolatti e C. Sinigaglia, *So quel che fai. Il cervello che agisce e i neuroni specchio*, Raffaello Cortina Editore, Milano, 2006.

34 Cf. S. Baron-Cohen*, Zero Degrees of Empathy: A new theory of human cruelty and kindness*, Penguin, Londra, 2025.

35 Cf. S. Gallagher, *Empathy, Silulation and Narrative*, in "Science in Context", Vol. 25, 2012, pp. 355-381.

This means, Gallagher continues, that empathic understanding of the other is always limited only to what the I has experienced first-hand, thus excluding from the simulation all situations outside its experience; how, then, would intersubjective openness to the other be explained if such understanding were limited only to what the I knows first-hand? How, then, would openness to the other be explained?

According to Gallagher, such an objection to ST renders the entire paradigm inadequate to explain the empathic phenomenon, also for the reason that it would always be limited to the experiential background of the empathizing subject; on the other hand, what could help explain empathy itself, according to Gallagher, is the observation that understanding the other is always based on one's own model of understanding and knowledge of the context in which the other acts and operates. In short, empathy, in order to exist, requires a wealth of narratives, what the author himself calls the "hermeneutic background of empathy," which provides the I with an understanding of the context in which the You acts and, consequently, what it might feel in that given situation. The ST lacks an explanation of how such a "context of understanding" or its role in the empathic phenomenon is obtained.

Although Gallagher's explanation focuses on a relevant aspect of empathy, namely the finding that there is an experiential framework that comes into play when "empathy is experienced," it fits into a heavily debated strand of studies on the "natural" relationship between empathy and altruism; the problem is to define whether the one is the prosocial behavior at the origin of the other, or vice versa. The aspect that is not shared here, because of the very intention of the thesis in question, is the starting assumption that the empathic phenomenon requires a narrative context already given to the I; this means accepting the indisputability of the phenomenon itself, leaving aside the strong conceptual limitations it presents, its fruition, for example, in subjects who are spatially and physically distant[36]; how, then, could we use empathy in order to "understand" the affective situation of the other when, in such a condition, we would not know the history or the context in which it is found? Moreover, the knowledge of the "already given" context calls to mind folk psychology, and thus does not move far from the structural problems that the latter poses.

It should be noted that the introduction of common sense psychology in the debate has led to a further change in the definition of "cognitive empathy" which, instead of focusing on the nature and epistemic justification of intersubjectivity, "reinvented" the problem by "including" it into the I itself: what kind of neurobiologically structured structure does the I possess to "enter into relationship" with the other? Or, rather, to "enter," through its thinking, into the mind of the You in order to know and interpret its intentions, beliefs, and desires in order to then "bring back into itself" the result of this shift outside the body, created by the mind of the I itself?

The problem regarding mentalization or mindreading took shape precisely starting from this question so generically summarized and, in this regard, the pillar of an entire debate is a target article by Peter Carruthers, *The relationship between mindreading and metacognition*[37] – published in 2009 – to which he is credited with both collecting and discussing the various arguments made in support of four generic models of explaining mentalization (*Model 1: Two mechanism, two modes of access; Model 2: One mechanism, two modes of access; Model 3: Metacognition is prior; Model 4: Mindreading is prior*) and with providing one that clarified the

36 Cf. M. Hoffman, *Empathy*, op.cit.

37 Cf. P. Carruthers, *How we know our minds: The relationship between mindreading and metacognition*, in "Behavioral and Brian Sciences", 32, 2009, pp. 121-138.

relationship between the mindreading structure directed toward the other and the "return into Self" of this interpretation.

The central point of Carruthers' position, which is helping to shift, once again, the focus of the debate on cognitive empathy, is the consideration that humans are "avid mind readers"[38]: they attempt to attribute to the other beliefs, thoughts, states or volitions experienced at that moment as an integral part of their own actions, triggering a process of "meta-representation", i.e., a process in which the object of study can become either the other than Self (therefore, we would be within the "classical" problem of "cognitive empathy") or toward Self; in the latter case a process of metacognition is set in motion, that is, the effort or process of thinking about the same "mind reading" thinking process, which results in an equal process of cognition, but directed toward oneself.

The questions that are following are inherent to the kind of relationship subsisting between the former and the latter process, posing the question of whether there is a difference (if so, what kind?) between access to other minds and access directed toward Self.

In this respect, in the mindreading/metacognition relationship several protagonists in the debate have accused Carruthers of eliminating the problem of the I's access to itself: this would no longer imply a relationship between reading the mental states of others and the underlying cognitive capacities, but a dualistic explanation of the relationship. How, in fact, is the I able to act, itself, as an object of study? And what "structures" would be involved: only the "neurobiological" ones, naturalizing the mental, or only those of the mind, distinct from those of the brain?

It is no coincidence that many advocates of the position advanced by Carruthers have suggested that he focus only on the I-other relationship. In fact, the way the U.S. philosopher's proposal has been structured would suggest a distinction between the mechanisms deputed to mindreading and those deputed to the interpretation of metacognitive states almost as if they were two distinct substances, one of which has already been inserted into the process and the other "to be inserted" at a later time[39].

In light of what has been said so far, it would be useful to include an outline of the steps elaborated and discussed in order to highlight the controversial aspects of the "empathy phenomenon" that would suggest conceptual limitations of both cognitive empathy and the mentalization process:

1. Transformation of the meaning of *sympathy* in Hume to "empathy," understood as *einpathos*, in both the artistic (Schleiermacher) and theoretical (Lipps) currents, assigning it the "ability to feel, first-hand, the emotions of others" and to provide knowledge;
2. Quine's revival of the concept in the "radical translation" experiment, assigning empathy an epistemological use;
3. Putnam and Nozick's specification of the characteristics of *simplicity and plausibility* and knowledge of others' behavior to the concept of empathy;
4. The risk of providing a metaphysical status to empathy and, thus, the fear of slipping into the dilemmas of Cartesian dualism, provide the *incipit* for seeking a transcendental and pragmatic nature to empathy; in particular, to the nature of the mental effort involved in the same capacity. Wilfrid Sellars, with his theorization

38 Id., *How we know our minds*, op. cit., p. 121.

39 Cf. D. Dennett, B. Huebner, *Banishing 'I' and 'we' from accounts of metacognition*, in "Behavioral and Brian Sciences", 32, 2009, pp. 148-149.

of the "Myth of Jones," provides the solution: the theorization of common sense psychology, namely, a "background theory already given and known," within which one objectifies the set of desires, beliefs and volitions that the You might feel (the mind of any member of the given community is objectified) and uses empathy as a mere search tool of the I for a suitable feeling in a given situation in the set of "already-given knowledge" provided by *folk psychology*;

5. The collapse of the two great families of common-sense psychology (the ST and TT) has shown its vulnerability, having given cognitive empathy only an intuitive explanation of its nature and not a theory based on its truth. In this regard, Premack and Woodruff's paper provides a solution, by further shifting the problem of the empathic nature: the inclusion of a theory of the mind in the explanation of cognitive empathy. In this way, the empathic capacity, from objectified tool in *folk psychology* "goes back to being" the "interiority" of the I in the theory of the mind;
6. The debate questions how, through a theory of the mind, the I is able to "understand" that of the You; Peter Carruthers' essay symbolically initiates the explanation of the process of mentalization, namely the ability to "read the mind" of others and "enter it" to understand their desires, beliefs and volitions. The division between basic empathy and cognitive empathy is betrayed and there is an attempt to explain the latter through recourse to the former.
7. The last frontier of totally sliding from the original intent of the meaning of *sympathy* occurs by the emergence of new questions concerning the "neurobiological capacity" of mindreading: how can we explain the kind of knowledge that the "mentalization module" can give the I if it were to address the capacity of mindreading to itself and no longer to the other?

This last question, it has been said, transforms the debate on cognitive empathy once again, raising the question of the kind of relationship there is between metacognition and mentalization, but on closer analysis it will become clear that this apparent novelty is actually the recovery of a phenomenon already theorized in the 1980s by John Flavell, metacognition which, in a forward-thinking manner, already explained decades in advance what the invention of mindreading denotes.

To this end, to suggest to the reader that such an interpretation will be necessary, a brief analysis of both what Flavell had meant by metacognition and reasons why it can be overlapping on the phenomenon of mindreading.

4. Metacognition and the overlap with the mentalization process

The discussion on metacognition in relation to mindreading is not an absolute novelty in the debate on metacognition itself, but it is in light of its transformation over time of the concept of cognitive empathy.

Upon a careful analysis of metacognition, both in its historical genealogy and in its ongoing development, we can see that as early as the 1960s it has made its way into the debate on emotions and the kind of knowledge they involve.

To be precise, U.S. psychologist Josef Hart[40] had identified the meaning of "metacognition" as a "feeling/perception of knowing"[41] in the light of his research, concerning the functioning

40 Cf. J. Hart, *Memory and the feeling-of-knowing experience*, in "Journal of Educational Psychology", 56, 1965, pp. 208-16.

41 In the current debate, this is the reason why we are focusing on the "tip of the tongue" phenomenon to demonstrate the critical role of metacognitive processes in the development of memory and, more generally, the issues involved in metacognition. For further discussion, see C. Heyes *et alii*, *Knowing Ourselves Together: The Cultural Origins of Metacognition*, in "Trends in Cognitive Sciences", Vol. 24, Issue 5, 2020, pp. 349-362; M. Rouault *et alii*, *Human*

of memory and how the person was able to monitor and control the set of information stored. He defined "metamemory" as the ability to establish the predictive reliability of the "feeling of knowing" in the evaluation of one's mnemonic abilities.

The term "meta," at least in Hart and in later studies inspired by him, referred to the subdivision of mnemonic ability: a first task related to remembering information and data; the second to monitoring and "calling to consciousness" such information, and perceiving this recollection through the "feeling of knowing".

While Hart is the symbolic leader of the debate on metacognition, influencing and stimulating different fields of research or investigation within it, we owe the most common and most discussed meaning of metacognition to an article published by John Flavell. As mentioned above, the American psychologist published a very short article in 1979, entitled *Metacognition and Cognitive Monitoring. A New Area of Cognitive-Developmental Inquiry*[42], in which, building on Hart's analysis, he argued that metacognition indicated knowledge and cognition about cognitive phenomena, that is, about all processes (cognitive, precisely) such as attention, memory, problem solving, self-control and self-awareness along with all social and cognitive processes of which an individual was capable.

The "field of investigation" of metacognition expanded considerably, no longer relegated only to the processes of memory control and monitoring, but to the totality of all the "knowledge" that the person could obtain, by turning their attention, memory, thinking and cognition, in general, toward themselves. From this point of view, the "meta" of the process identified by Flavell no longer stopped at the mere distinction of roles and capacities of mnemonic tasks – as in Hart – but indicated the possession of a theoretical knowledge of what the subject knows, understands, and remembers and the related processes (what, when and why). Therefore, a kind of "subdivision of levels" between the thought of an I, looking outward or towards a You, and a more complex thought looking toward oneself; what many years later would become "mind reading".

In the first few paragraphs of the article, the American psychologist endorses the difference between *metacognitive knowledge*, which, as noted, is the result of the process of self-questioning by the Self, and *metacognitive* experience, overlapping with Hart's "feeling of knowing," arguing that both can be considered almost the same phenomenon: one influences the other and vice versa, precisely because the activation of purposes or strategies of action for the achievement of an end (the other two classes of the phenomena he identified) affect both access to memory and the awareness of one's cognition: metacognitive experience is the former; metacognitive knowledge is the latter. Therefore, since metacognitive knowledge has no different nature than other knowledge "stored" in one's memory, according to Flavell it can also be defined as such when it influences the course of action taken by an agent for a purpose in an unconscious manner; this means that awareness or consciousness does not play a relevant role in "self-mastery" and that it "simply" represents an option.

Beyond this controversial aspect, it should be noted that if we were to provide a merely phenomenological explanation of metacognition, it could be outlined in this way:

- X conceptually represents mental states;
- X conceptually represents [others', Ed.] mental states with their intentional content toward which they are "directed" [...];

Metacognition Across Domains: Insights from Individual Differences and Neuroimaging, in "Personality Neuroscience", Vol. 1: e17, 2018.

42 J. H. Flavell, *Metacognition and Cognitive Monitoring. A new Area of Cognitive-Developmental Inquiry*, in "American Psychologist", Vol. 34, No. 10, 1979, pp. 906-911.

- X understands, through certain meanings, the relationship between an agent's mental states and their respective behavior [...]. This understanding enables X to make predictions about others' behaviors and to explain their behaviors[43].

Paraphrasing the above, in the first aspect we could refer either to cognitive knowledge or the metacognitive experience: an agent may or may not be aware of their cognition and have first-hand experience of it; in the second, by using the three sub-categories identified by Flavell as pertinent to the nature of metacognitive knowledge (people, tasks or strategies), it is explained that in the inter-relationship with the other, the cognitive effort to "understand" the You does not identify a specific ability to go beyond oneself to "simulate," "reproduce," or "change one's point of view" to produce knowledge about the other, but a self-directed ability to direct the "curiosity to understand oneself" toward oneself; in the third passage, Flavell, referring to the use of the "strategy" subcategory of metacognitive knowledge, claims that the latter represents the first step in enacting the agent's manifest behavior in the world. An emblematic example might be that of a little girl who, in order to relate to her little brother, knows that she will have to adopt "strategy A" in "task X" instead of "strategy B" for "task Y", implicitly suggesting that metacognition has its own relational and intersubjective openness[44].

If this clarification convinces as to the correctness of the phenomenological explanation of metacognition, one should not be surprised if the above schematization of the various steps of metacognition, explained through the use of Flavell's thesis, is in fact the phenomenological schematization of the three key steps of the process of mentalization used by Daniel Hutto, as it is presented in the debate today.

This would thus suggest that mindreading (the latest transformation of the concept of "cognitive empathy") and metacognition are the same phenomenon, at least as far as the phenomenological aspect is concerned, and that speaking of "cognitive empathy" generates conceptual errors and misunderstandings about the way intersubjectivity between an I and a You occurs.

5. The naturalization of the I and the "relational void"

The course outlined so far has aimed to highlight the strong conceptual limitations of cognitive empathy and its possible overlap with the concept of metacognition; in order to succeed in this endeavor several key steps have been taken which, chronologically, starting from the transformation of the concept of *sympathy* in Hume have reached the meaning of "entering into the emotions of others, experiencing the same intensity"; from the objectification and abandonment of the concept of mind by *folk psychology* to its representation in the debate by Premack and Woodruff's paper; from the neurobiological ability to explain the cognitive effort of empathy through the "mindreading module" to the relationship between mentalization and metacognition, generating a further shift in the meaning of cognitive empathy.

It is difficult to find a philosophically convincing justification for the latter step: the possible overlap between mentalization and metacognition suggests that they are not

43 D. D. Hutto, *ET ALII*, *Editorial: Social Cognition: Mindreading and Alternatives*, op. cit., p. 376.
We read: *The standard view is that an agent X engages in mindreading only if:*
1. *X conceptually represents mental states (e.g., beliefs, desires and perceptions).*
2. *X represents mental states with their intentional (with a "t") content, i.e., that which they are "about" or "directed toward." Mental contents are typically assumed to be propositions specifiable by "that" clauses-for example, someone might believe that "the food is located under the bucket."*
X understands, by some means, the relations between an agent's mental states, their environmental conditions, and their behavior. This understanding enables X to make predictions about others' behaviors and to explain those behaviors.
44 H. Flavell, *Metacognition and Cognitive Monitoring*, op. cit.

two different and distant phenomena, but one and the same, interpreted from different perspectives.

Therefore, if cognitive empathy/mentalization does not seem to be a satisfactory explanation of intersubjectivity, it remains to be explained how an I can relate to a You and how it can "understand", as if it were the active protagonist of this emotional narrative, the emotions felt by the other.

An alternative explanation would be to look at the inter-relationship in a different way, starting precisely from the nature of the I or the person acting toward the other's subjectivity.

While fully sharing the program of cognitive naturalization[45] and cognitivist research in the philosophy of mind on the fragility and unreliability of the I[46], one is inclined to argue that the greatest deception to which the sense of one's own subjectivity falls victim is precisely the illusion of a simplified image of an "acting, conscious, and aware Self," constructed by the brain as the result of intense brain activity. Therefore, at the basis of intersubjectivity, there is not the (conscious, rational, in the first person) I acting toward the You, but an intense neural activity, defined as the *Self* which, to survive, gives itself an "ideal" image (the I) with which it can relate to the context. Interestingly, in recent years, the naturalized approach to the mental has been extended to an evolutionary explanation of the mental itself: taking up "Darwin's law" on the "survival of the fittest" and on genetic mutation as the basis for the structural change in the species, the "mind phenomenon" is included within this explanatory logic; thus considered as the ultimate result of a selection process that has led from the appearance of bacteria to authors such as Johann Sebastian Bach[47]. In the wake of what has just been said, Daniel Dennett, for example, also considers the mental experience the result of an illusion arising from complex brain mechanisms that provide first-hand experience of "being a *cogito*" and "not just a body". A useful and functional stratagem, evolutionary in nature, for the survival and adaptation of a species in a given context, which would ward off dualistic explanations of the mental[48].

The first major transformation that these approaches entail in the debate on intersubjectivity is the change in the nature of the protagonists involved: no longer an "I" relating to a "You", but a *Self* relating to a You, signifying that in the above-mentioned relationship there would never be an "underlying sincerity" that guarantees that both participants understand each other's emotional states. Hence, this gives rise to the consequence on the intersubjective relationship, which completely changes its meaning, resulting in the existence of a double "void" to be filled in the *space of the relationship* between a *Self* and a You, since:

- On the *Self's* side, personal agentivity is constructed after the action of inter-relation has been implemented and, illusively, the *Self's* "consciousness" has self-narrated that it has been the author of the action "all along" (consciously, the *Self* has established a relationship with the You).
- On the You's side, in addition to the same mechanism experienced first-hand by the You itself, there is also the illusion of having communicated its desires and beliefs to the other *Self* just as the You experiences them first-hand.

45 Cf. S. Nannini, *Naturalismo Cognitivo. Per una teoria materialistica della mente*, Quodlibet Studio, Macerata, 2007.

46 Cf. D.Wegner, *Who is the Controller of Controlled Processes*, in R. R. Hassin, J. S. Uleman, e J. A. Bargh (a cura di), *The New Unconscious*, Oxford University Press, Oxford (UK), 2005; Id., *The illusion of Conscious Will – New Edition*, The MIT Press, Cambridge (Massachusetts), 2018; M. Gazzaniga, *Who's in Charge? Free Will and the Science of the Brain,* Robinson, London, 2016.

47 Cf. D. C. Dennett, *From Bacteria to Bach and Back: The Evolution of Minds,* W.W. Norton and co. Inc., New York, 2017.

48 Cf. K. Frankish, *What is illusionism?*, in "Klēsis Revue Philosophique", n. 55, 2023, pp. 1-13.

This explanation is what is meant by "relational void"; the result of a structural incommunicability between interacting subjects as a result of the functional deception that the brain performs to itself as the structuring of a self-image, projected outwardly: the "I."

From this it follows that there will always be a double void to be filled, both in the acting subject and in his interlocutor; in a *Self*-You type scheme, the first "void" is created between the *Self* and the You in the space of relationship: how could the *Self* be "certain" that what it will say or do will be the original result of its intentions and not an explanation that arose only after the action has already been enacted?

The second "void" is created in the communication from the You to the *Self* of one's intentions, beliefs, and desires: how could the You be certain that what he himself wants to communicate is understood in the terms in which he himself communicates them to the *Self*?

The result of the "relational void" is the structural incommunicability between any interacting subject; an inability, therefore, to "empathize with the other," initiating a cognitive effort such as to arrive at the "reading the mind" of the other, since both the *Self* and the You are unable to "structurally step outside themselves" to understand intentions or behaviors that are carried out from a third-person perspective. In other words, neither will the *Self* be able to "come out of itself" to transfer, transport and, thus, understand its own point of view in that of the other; nor will the You ever be certain that the other has initially understood what it is communicating.

The reference to the *monad* would seem almost natural; however, the difference between Leibnitz's conception and that of the "relational void" is evident in the "solution" provided: while for the philosopher from Leipzig the inter-relation between two or more monads took place through reference to the *pre-established harmony*, in this position intersubjectivity is possible through metacognition first; then through the shifting of one's beliefs, volitions and desires onto the other, by means of the imagination.

Therefore, one finds oneself in a state of a complete absence of relation and intersubjectivity which, however, seems in fact to be denied by the daily social relationships that everyone has: how can we explain this apparent contradiction?

The phenomenological process underlying "relational void" is able to overcome and resolve such apparent incompatibility precisely through the use of metacognition.

In fact, the *Self*, thinking that it "authentically" understands the account of the You, "narrates" itself to find itself in these same conditions; for the *Self*, the result would be to imagine the emotional sphere it would feel if it were in that hypothetical context and the relative possibility of narrating what it feels illusively. Only at this point, the use of metacognition having ceased, will the *Self* be able to "transfer" such an imaginary experience onto the You.

In this context, incommunicability derives as much from the sender, the You, as from the receiver, the *Self*, since both are unable to "come out of themselves"; the debate on cognitive empathy thus seems to have focused on the result of the complex naturalized system of encephalic activity, resorting to "mysterious" capacities or tools (empathy, cognitive empathy, mindreading) to explain intersubjectivity, emphasizing only the "immersion in feeling" and ignoring the causal process that generated it.

In the light of this explanation, the "relational void" could be outlined as follows:

1. The *Self* assumes that it has *understood* a situation of distress communicated by the You;
2. The *Self* imagines its own agentivity in the same experiential framework;
3. The *Self* understands what emotions it might experience in the same situation;

4. The *Self* transfers the hypothetical life experience onto the You, narrating to itself the hypothetical emotional sphere it might experience;
5. The *Self* relates to the You, telling it that it has "understood" the situation of distress, thinking, illusively, that it has "entered into the mind" of the You and experiencing first-hand the *same* emotional sphere as the You.

Behind such "apparent simplicity" lies a complex *mental* operation capable of providing everyone with the illusion of understanding, making oneself understood and relating to the other; once again, a "magnificent trick" that the brain "plays" on itself.

Therefore, in the light of what has been argued so far, we would like to support the abandonment of the concept of cognitive empathy and its derivatives in order to opt for a *simpler and more plausible* explanation of intersubjectivity.

Conclusions

In light of what has been discussed, this paper did not aim to conclusively demonstrate the illusory nature of cognitive empathy, but rather to raise a number of critical arguments against some of its current interpretations. The reference to the key steps in the emergence and transformation of the concept of cognitive empathy suggests that there has been a misunderstanding of the use of the term, its concept and its aims in the evolutionary history of the concept; indeed, from Hume to the present day, the very meaning of empathy has changed, from sympathy as the origin of morals, according to the Scottish philosopher, to the reading of the mind in the current debate. One possible interpretation of this change could be advanced by suggesting that the naturalization of the Cartesian *cogito* may have provided a practical element to explain both the nature of the mental and intersubjectivity. This interpretation finds its objective confirmation in the theorization of folk psychology through the 'Jones myth' and its subsequent abandonment when the two great families of common sense psychology (the ST and the TT) encountered considerable difficulties in guaranteeing an explanation of the mind consistent with the premises from which they both proceeded.

This constant shift in meaning, and the departure from the very sense in which Hume hypothesized *sympathy*, may explain why, to this day, cognitive empathy seems to be a meaningless concept. What remains constant, however, is the explanation that one can "experience the other's emotions first-hand", as if one could actually "enter the other's mind" through mindreading in order to gain knowledge of it. It is no coincidence that, in the light of an emerging literature, attempts are being made to legitimize the process of mindreading by explaining it in terms of its relationship to metacognition.

With this in mind, an attempt has been made to show that when cognitive empathy is read through the lens of metacognition itself, in the context of the naturalization of the mental, it runs the risk of appearing as a redundant, if not misleading, concept. Indeed, by suggesting the naturalization of the mental, emphasis has been placed on the consequent illusory nature of the I, that is, the one who, it is assumed, initiates a relationship with the other, feeling "empathy" for what the You itself communicates to it. This would entail an initial structural transformation of the intersubjective relationship: it is no longer the I that relates to the You, but the *Self* that relates to the You, pointing out that the Self identifies the complex brain activity at the origin of the illusory nature of consciousness (the I) as the "unity of oneself".

The consequence of this clarification would be what has been called "relational void": the existence of a temporal void in the relationship between cerebral activity and the consequent unified image of the I, which leads to the illusion of having first consciously wished to establish a relationship with the You and, subsequently, having brought it into being. With regard to the You, in addition to the same illusion, we also add that of believing that one has "authentically" communicated the emotions of one's experience to the interlocutor; a false

belief dictated by the structural incommunicability of each one, of each *monad*, to initiate any relationship, since it is not able to "come out of itself".

While the "relational void" proposed in the concluding pages is only a theoretical sketch, it is intended to present a possible alternative for rethinking the relationship between I and You, beyond the classical structures of simulation or the theory of mind.

While taking a critical stance towards some formulations of cognitive empathy, this paper does not aim to exhaust the issue or to deny the possibility that other notions – for example, those inspired by the phenomenology of perception or the 4E models – might provide a more adequate account of intersubjectivity. Rather, it aims to circumscribe and challenge a conceptually fragile use of cognitive empathy in the hope of stimulating a more analytical discussion of the theoretical assumptions that underpin its use.

REFERENCES

Almeida, R., & alii (2024). Neural Correlates of empathy: a systematic review of event-related potentials studies in perceptual tasks. *Brain. Sci*, 14(5), 504;

Aaltola, E. (2014). Affective empathy as core moral agency: psychopathy, autism and reason revisited, *Philosophical Explorations*, 17:1, pp. 76-92;

Baron-cohen, S. (2025). *Zero Degrees of Empathy: A new theory of human cruelty and kindness*, Londra: Penguin;

Cameron, C.D., Conway, P., & Scheffer, J. A. (2022). Empathy regulation, prosociality, and moral judgment, *Curr Opin Psychol*., Vol. 44, 188-195;

Cappiello, I. (1983). La morale della simpatia in David Hume; Napoli: Liguori Editore;

Carruthers, P. (2009). How we know our minds: The relationship between mindreading and metacognition. *Behavioral and Brian Sciences*, 32, 121-138;

Cuff, B., Brown, S. J., Taylor, L. & Howat, D. (2016). Empathy: a review of the concept. *Emotion Review*. Vol. 8, Issue 2, 144-153;

Decety, J. & Ickes, W. (2009). *The Social Neuroscience of Empathy*. Cambridge (MA): MIT Press;

Decety, J. & Ickes, W. (2011). *These Things Called Empathy: Eight Related but Distinct Phenomena,* from *The Social Neuroscience of Empathy*. Cambridge: MIT Press;

Dennett, D. C. (2017). From Bacteria to Back and Back: The Evolution of Minds, New York: W.W. Norton and Co. Inc. ;

Dennett, D. & Huebner, B. (2009). Banishing 'I' and 'we' from accounts of metacognition. *Behavioral and Brian Sciences*, 32, 148-149;

De vignemont, F., & Singer, T. (2006). The empathic brain: How, When and why?. *Trends and Cognitive Sciences*, Vol. 10, n. 10, 435-441;

Eisenberg, N. (2018). Empathy and Sympathy, in M. Lewis and j. M. Haviland-jones [eds], *Handbook of Emotions* [pp. 677-691]. New York: Guilford Press;

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring. A new Area of Cognitive-Developmental Inquiry. *American Psychologist*. Vol. 34, No. 10, 906-911;

Frankish, K. (2023). What is illusionism?. *Klēsis Revue Philosophique*, n. 55, 1-13;

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17, 535-554;

Gallagher, S. (2012), Empathy, Silulation and Narrative. *Science in Context*, Vol. 25, 355-381;

Gazzaniga, M. (2016), *Who's in Charge? Free Will and the Science of the Brain*. Londra: Robinson;

Hart, J. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56, 208-16;

Heyes C. & alii (2020), Knowing Ourselves Together: The Cultural Origins of Metacognition. *Trends in Cognitive Sciences*, Vol. 24, Issue 5, 349-362;

Hoffman, M. (2010). *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge: Cambridge University Press;
Hume, D. (2023). *An Enquiry Concerning the Principles of Morals*. Firenze: Paperback;
Hume, D. (2023). *A Treatise of Human Nature*. Firenze: Paperback;
Hume, D., & alii (2011). Editorial: Social Cognition: Mindreading and Alternatives. *Rev. Phil. Psych*., 2 309-315;
P. Jacob (2011). The Direct-Perception Model of Empathy: A Critique. *Review of Philosophy and Psychology*, 2, no. 3, 519–540;
Krebs, D. L. (1975). Empathy and Altruism. *Journal of personality and social psychology* XXXII, 1134-1146;
Krueger, J. (2018), Direct social perception, in A. Newen, L. De Bruin, S. Gallagher [eds]. *The Oxford Handbook of 4e Cognition*, pp 301–320, Oxford: Oxford University Press;
Lewis, D. (1972). Psychological and theoretical identification. *Australian Journal of Philosophy*, n. 50;
Lewis, M., & Haviland-Jones, J. M. (2018). *Handbook of Emotions*. New York: Guilford Press;
Nannini, S. (2007). *Naturalismo Cognitivo. Per una teoria materialistica della mente*. Macerata: Quodlibet Studio;
Premack, D. & Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind?. *Behavioral and Brain Sciences*, n. 4, 515-629;
Putnam, H. (1995). Mind, Language and Reality. *Philosophical Papers*, Vol. 2, 70-84;
Rizzolatti, G. & Sinigaglia, C. (2006). *So quel che fai. Il cervello che agisce e i neuroni specchio*. Milano: Raffaello Cortina Editore;
Rouault, M. & alii (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, Vol. 1, 1-17;
Stietz, J., Jauk, E., Krach, S., & Kanske, P. (2019). Dissociating Empathy From Perspective-Taking: Evidence From Intra- and Inter-Individual Differences Research. *Front. Psychiatry*, 10:126;
Stueber, K. R., *Rediscovering Empathy: Agency, Folk Psychology and the Human Sciences*, MIT Press, Cambridge, 2010;
Wegner, D. (2018). *The illusion of Conscious Will – New Edition*. Cambridge (Massachusetts): The MIT Press. ;
Wegner, D. (2005). Who is the Controller of Controlled Processes, in R. R. Hassin, J. S. Uleman, e J. A. Bargh [eds], *The New Unconscious*, Oxford: Oxford University Press;
Wispè, L. (1986). The distinction between sympathy and empathy. *Journal of personality and social psychology*, Vol. 50, n. 2, 314-321.