

Understanding empathy: Metaphysical starting assumptions in the modeling of empathy and emotions

Joel Parthemore ¹

Abstract. This paper has three main purposes: to set out the relationship between empathy and related phenomena, including emotional contagion; to explain how metaphysical starting assumptions regarding the nature of empathy predispose one toward one or another account of these phenomena and toward different interpretations of the same empirical data -- often radically different; and to use recent discussions of empathy in the phenomenological and enactive communities (in particular their rejection of *theory of mind* accounts) to put forward a radical proposal. In the paradigmatic cases, one feels that one is feeling (at least some substantive portion of) what another person is feeling: “I feel your pain”. But there are certain intense experiences along with certain related but less intense ones where there is, I claim, a single joint experience among two or more individuals. One could call these experiences “*extreme*” empathy. This is how phenomenologists should, I think, cash out the frequent claim that in many circumstances, one agent “directly” experiences the emotional state of another without requiring the mediation of anything like theory of mind.

1 INTRODUCTION

Empathy I will take to be the feeling that one is experiencing, to some substantial extent, what another is feeling – often facilitated by some corresponding experience(s) in one’s past but without requirement to be in precisely the same emotional state.² How one understands it has a great deal to say about what are often unstated assumptions on the nature of personhood: one of the key starting points from which we engage in theoretical and computational modeling of affect and emotion in general, and emotional states like empathy in particular.

Sympathy by contrast merely entails showing awareness of and consideration for the mental/emotional states of another (particularly when the person is going through strongly negatively valenced times). It is often understood pejoratively, as a kind of poor substitute for empathy: “I sympathize, but I do not empathize”.

The question at hand in this paper is, must empathy be accounted for in terms of so-called *theory of mind*, and are breakdowns in empathy to be explained in terms of a theory-

of-mind deficiency? If not, what is the alternative – or, if more than one, what are the alternatives?

1.1 Metaphysical starting assumptions

Whether any of the standard positions in this area is ultimately “right” – if that notion is even coherent – is ultimately and almost certainly a metaphysical, not an empirical, question. That is, it concerns initial axioms rather than observations. Like any metaphysical claims, the value lies not in knock-down proof but explanatory value: if one assumes *these* things, then *this* is what follows. The consequent explanations can be judged superior or inferior on the basis of seeming simplicity, completeness, and consistency. Nevertheless, it will still be the case at the end of the day that one could make a valid case for any of them – with the caveat that some may prove easier to defend than others³ and the likelihood of other alternatives that haven’t received due consideration or yet been considered at all.

My interests here are pragmatic and (despite the possible tone otherwise) pluralist, likewise more epistemological than ontological. I am less concerned with the “actual” underlying reality – insofar as it is knowable (my preference is toward a Kantian view) – and more with how we currently do, and how we most usefully can, model these things conceptually, as lay persons and as researchers. Doing so requires, I think, combining “top down” with “bottom up” approaches to mind and cognition, and I continue to be mystified by those whose focus seems exclusively on one or the other, as one finds with certain “traditional” AI accounts on the one hand and some connectionist/associationist approaches on the other.

Of course it is generally more practical to focus at any given time on one or the other, as my own (limited) empirical work has done. The mind-mapping software I wrote for my doctoral thesis [11] as a direct translation of its theoretical content into software form – specifically, of the Unified Conceptual Space Theory [13] – very much takes a top-down approach in offering

¹ University of Skövde, Sweden, email: joel.parthemore@his.se

² ... A point that Dan Zahavi [17] is keen to stress. Likewise he challenges those accounts of empathy that attempt to connect it to “mind reading” or telepathy [17, p. 295].

³ The principle applies much more broadly. In principle, one could defend an updated version of Tycho Brahe’s epicycle-based model of the solar system taking into account all of our subsequent explorations and observations. However, the consequent model would be so complex as to put most people off even attempting it. Most of the time people argue endlessly and, one might suspect, fruitlessly over whether and how Ockham’s razor applies, not least because there is no untendentious metric for simplicity/complexity; it is, perhaps, only in the extreme cases that its application is clear, if only for practical necessity.

a specific algorithm that all concepts, regardless of what they are concepts of⁴, are meant to follow. On the other hand, there seems no reason in principle why the same algorithm could not be used in a bottom-up, exploratory manner by, say, embedding it in a robot. By contrast, the SEER-3 robotic platform [2], which I developed in collaboration with Ron Chrisley to model the non-conceptual content of visual experience, works bottom up: the robot looks around and gradually assembles a 2D model of its environment (albeit with a very much top-down model imposed on it for foveal/non-foveal/peripheral vision, memory fading, and motion detection, never mind of what counts as “nonconceptual content of visual experience” in the first place).

Like everyone, I do have ontological commitments, of course: first of all, that emotions are evolutionarily ancient and, as such, intimately tied up with our existence as living organisms and our nature as that small subclass of living agents who comprise rational conscious agents. They can be and often are in tension with “high-level” reasoning but should in no way, pace a number of familiar accounts from the rationalist tradition (and sympathetic voices from symbolic AI), be seen as antagonistic toward it.

1.2 Empathy and its kin

To understand empathy, it is important to separate it from related phenomena including sympathy, emotional contagion, and emotional sharing⁵ and critically examine those cases where empathy is said to be broken or missing altogether: i.e., sociopathy/psychopathy and Asperger’s/autism. My core claim is for an empathy-related phenomenon – call it “extreme” empathy – best understood in terms of a single conscious experience involving two or more agents. This in turn opens the way for the possibility that (at least as a passing state, but conceivably on a more stable basis as well) conscious agents are individuals or parts of individuals on more than one level.

Emotional contagion involves “catching” an emotional state analogously to catching a cold. The “infected” person(s) is under no requirement to understand what it is all about. The source for the initial “infection” is irrelevant, for the emotional state takes on a life of its own: e.g., one person starts laughing for the sole reason that someone else is laughing; if there are more people in the room, pretty soon everyone may be laughing.

Emotional sharing involves being in the same context perturbed by the same motivating forces toward the “same” emotional states. Again, the individual source of the emotional state – if there even is one to identify – is not relevant; one even supposes emotional sharing to originate at the group level. Stereotypical occasions include weddings, funerals, and sporting events.

Empathy is different these phenomena in several key respects:

- The source of the emotional state crucially *does* matter.
- A critical emotional connection is made between the source and the one empathizing (note that the empathic relation is asymmetrical).

⁴ . . . Including, of course, concepts of emotions.

⁵ I owe much of this discussion to [17].

- The one empathizing must feel some substantial part of what the source is feeling emotionally, even if she is not ultimately in the same emotional state.

1.3 Breakdowns in empathy

Supposed absences of empathy resulting from mental disorder should be examined critically. Sociopaths and psychopaths⁶ are often cited as lacking empathy. Research reported in 2013 [6] suggests that, rather than lacking empathy, these people can switch it on and off at will; the same research raises the intriguing notion that empathy is both trainable and mis-trainable. Furthermore, one cannot discount the possibility that they have well-developed *involuntary* empathic responses but that those responses are not the “right” (i.e., socially accepted/acceptable) ones: i.e., they feel some substantial part of what the other person is feeling emotionally (recalling my opening definition), but their further emotional response suggests a kind of empathic misalignment.

The other group often cited for empathic deficiencies, according to the “standard” position articulated most clearly by Simon Baron-Cohen, are those diagnosed with Asperger’s syndrome or autism spectrum disorder.⁷ As with the sociopaths/psychopaths, they appear to be poorly attuned to many if not most social cues. Again, however, recent research (see e.g. [9]) suggests that Asperger’s persons and autists may not be insensitive to others’ emotional states but instead *overly* sensitive to them. The metaphor is of someone shouting continuously in one’s ears; one places one’s hands over one’s ears to dampen the sound.

1.4 The remainder of this paper

Section Two examines the “traditional” account of empathy and the phenomenologists’ response, which I take to be essentially in line with that of the enactivists. Section Three then explains where, I think, the phenomenologists go wrong, and why one needs to discuss another related phenomenon that resembles empathy but that goes, in significant ways, beyond it: i.e., “extreme” empathy.

2 THE “TRADITIONAL” ACCOUNT AND ITS DETRACTORS

The theory-theory of mind and the simulation theory of mind both deny that it is possible to experience other minds, both presuppose the fundamental opacity or invisibility of other minds. . . . Both accounts consequently share the view that the minds of others are concealed and hidden, and they consider one of the main challenges facing a theory of social cognition to be the question of how and why we ascribe such hidden mental entities or processes to certain publicly observable bodies. [17, p. 286]

⁶ I will not attempt to address the question of whether sociopathy and psychopathy describe distinct phenomena or a single phenomenon.

⁷ The *DSM-V*, of course, combines these into a single diagnosis. Note that enactivists are inclined to view autists as exhibiting cognitive diversity rather than intrinsic disorder; see [12].

One of the key debates, when it comes both to empathy and the wider field of emotion modeling, lies with the role of so-called *theory of mind*, where “theory of mind” is the capacity to attribute beliefs, desires, intentions, motivations, knowledge, etc. to oneself and others – correctly, most of the time – and to understand and anticipate that the perspective of others will often differ from one’s own. It is attributed to young children through the so-called *false-belief task*, which most children are able to pass by the age of four, though children diagnosed autistic are generally only able to do so much later (in cases of “low-functioning” autism, perhaps not at all). The implication is that one understands others through being able to theorize – to form theories – about their minds. Proponents of what has been called theory theory take that quite literally; simulationists do not.⁸

Phenomenologists, including Dan Zahavi, Joel Krueger, and others, like to talk about the “direct” experience of another’s mental/emotional states, unmediated by any so-called theory of mind. They wish to reject both the theory-theory and simulationist accounts with their common assumptions that the experience of conscious agents is opaque to all but themselves and requires some kind of inferential leap to grasp. They are not rejecting theory of mind outright (at least, most of the time) but assigning it to a far more peripheral as opposed to foundational role – where understanding others emotionally and cognitively does not proceed from theory of mind, for the most part, but *vice versa*. If that is the negative claim, then the positive claim is that one finds experience of the cognitive states of others that is direct, unmediated, and non-inferential. It constitutes a perceptual modality, both like and unlike other modalities. Most critically for my purposes though, it generally (and ironically) retains a key assumption with the theory-theory and simulationist accounts: namely, that the nature of the individual is essentially isolated and (fully) distinct from other individuals. In place of opacity one finds, at most, occasional transparency, and otherwise translucency.

The rejection of theory-of-mind language – except, perhaps, with respect to certain narrow and well-defined applications (when we *are* theorizing about what and how other people are thinking) – is, I think, correct. Otherwise – echoing a familiar point from the enactivists – such language risks greatly over-intellectualizing processes that are largely intuitively, emotionally, experientially, and pre-intellectually driven. We don’t come to understand how other minds work by forming theories and testing hypotheses in some empirical-science-like way; rather, we come to understand other minds by learning to see the “us in them” and the “they in us”, and any theorizing comes only late in the day. We commonly think of other people being like ourselves, and ourselves as being like other people – with e.g. similar beliefs and motivations -- until established otherwise. In other words, we come to understand other people through understanding ourselves and ourselves through understanding other people.

The retention of the isolated consciousness – a point which I have discussed in person with Zahavi – is where I part company. The notion of “direct” experience is key. If one accepts that talk of “direct” experience of another’s state of mind is meant to be literal, not metaphorical, then how should one

cash it out? Here, the phenomenologists grow suddenly very cautious, if not in fact reticent.

3 BEYOND EMPATHY

I suggest that phenomenologists borrow a page from Andy Clark and Dave Chalmers’ *Extended Mind Hypothesis*, with its provocative opening gambit: “where does the mind stop and the rest of the world begin?” [5, p. 7], along with Clark’s subsequent discussion of “profoundly embodied agents. . . able constantly the negotiate and renegotiate the agent-world boundary itself” [3, p. 34] – a world that is simultaneously a physical world and a social one (physically realized, but conceptually distinct and irreducible). Enactive philosophers – in the tradition of Humberto Maturana, Francisco Varela, Evan Thompson, Dan Hutto and others – tend to see the Extended Mind Hypothesis as too conservative, leaving much if not most of cognition “inside the head”, with only occasional elements leaking out into the surrounding social and physical environment. Cognition – and therefore mind – is not, they would say, in the world or in the agent (they would say this creates a false dichotomy) but rather, so far as it is located anywhere, “in” the interaction between agent and environment. Rather than world creating agent or agent creating world, agent and world are co-created through interaction, each bringing the other forth⁹; and agent is ultimately continuous with environment. In this way enactivists reject both internalist and externalist views on cognition. They leaves them open to the possibility of not just extended mind but extended consciousness: something that Chalmers has shown explicit willingness to entertain¹⁰ and Clark has done his best to reject [4] (on the argument that the available “bandwidth” is too “narrow”).

What I think that phenomenologists should say is that consciousness is only ever relatively private and never absolutely so; the view of human beings as isolated islands of consciousness is a modern Western conceit (rather than unassailable statement of fact or cultural/historical universal) and should be recognized as such. Agents, as biological organisms, have reasonably well-defined physical boundaries. Agents, as conscious cognitive entities, do not. (Yet these are not two different things – two different agents somehow! – but rather two competing, complementary, yet ultimately irreconcilable perspectives on one and the same agent in one and the same world: a version of dual-aspect monism that I have referred to as *perspectival dualism* (see e.g. [15]), to stress that it is, in fact, a form of dualism (or rather, both dualism and monism). To say that minds are physically instantiated – as surely they are – is not necessarily to accept that they are neatly physically localized, as is all too commonly assumed – in which case, they will have neither an “inside” nor an “outside” except insofar as is convenient to imagine them doing so.

In a recent paper [14], I raised the possibility of a new theory of personhood that would do away with the common – but far from universally accepted (consider the perspective from a number of Eastern cultures) – view of conscious minds as strictly isolated islands. Rather, I suggested there, conscious-

⁹ “I have proposed using the term *enactive* to. . . evoke the idea that what is known is brought forth, in contraposition to the more classical views of either cognitivism or connectionism” [10, p. 255]

¹⁰ See his introduction to [3, p. xiv].

⁸ cf. [17, pp. 285-286].

ness is only ever relatively private and never absolutely so. As extended mind suggests, there is no clear line not only where mind stops and world begins, but likewise where one mind stops and the next begins. On such a view, minds “bleed over” substantively into one another; people both are substantively defined by, and help to define, the people around them. Change the community, change the social environment, and – in important ways – one changes who the person is. Consider the “brain-bound” mind as the centre of a circle that expands and contracts in changing circumstances.

So the way I think phenomenologists should cash out talk of “direct experience” is not through talk of empathy but rather by going beyond empathy and taking a much more radical stance that I have been hinting at throughout this paper: namely, that the line where one person stops and the next begins is a flexible, pragmatic one and that, in certain contexts, that line drops out altogether. In such moments, there is no “you” nor “I”, but only a “we”. There are not two distinct experiences aligning with each other through some empathic mechanism but one unified experience (albeit one that may well start from empathy). Sometimes, when we see another person crying or laughing – experiencing some intense emotion – we do not feel that experience as our own. Rather, there is a common experience distributed among two (or sometimes more) people. For that moment at least, there are not two individuals, but one: for a single joint experience implies (for the duration of that experience) a single individual. The upshot is that individuality is not a fixed quantity, and that – at least part of the time – we constitute not individuals in our own right (or not *only* individuals in our own right) but rather parts of larger, collective individuals.

Of all phenomenologists with whom I am familiar, Krueger comes closest in print, I think, to allowing such a possibility – particularly with his talk of certain emotional states (the paradigmatic example for him involving perhaps infant and caregiver) where the minimal unit is a dyad, not an individual. Consider: “from birth, some experiences are constitutively [non-reductively] social in that certain phenomenological states, such as the positive emotions that arise within these early exchanges, are jointly owned” [7, p. 509]. The infant learns to experience certain feelings by first having them as joint experiences with the caregiver. I would like however to make explicit what I take Krueger to leave implicit: that the joint experiences involve one unified experience for two (or more) people; and furthermore allow that such shared states extend throughout our lifetimes – albeit by becoming, over time, more often than not the exception rather than the norm. This includes many states that currently – and, I think, mistakenly – are understood in terms of “ordinary” empathy.¹¹

What circumstances do I have in mind? One thinks of certain sexual encounters, where there is the experience of one individual in two bodies; certain intense engaged conversations that mirror extended introspective consideration of some subject or another; certain other moments of joint trauma, or of joint pleasure or euphoria (reminiscent of Krueger’s early bonding experiences), where the conventional lines of self and other get blurred; certain other situations of intense engage-

ment with another person of a calmer and more extended nature¹²: where e.g. each member of a long-term couple reliably completes sentences for the other.

4 CONCLUSIONS

However it best should be understood, the dividing line between agent and environment, self and other cannot be taken for granted as fixed and determinate. When Robert Rupert [16] and Frederick Adams and Kenneth Aizawa [1] claim that all the available empirical evidence points to mind (and therefore consciousness) being limited to the boundaries of brain and skull, they are – it seems to me – begging the question, precisely by what they accept for evidence. At the same time, as others have argued (see e.g. [8]), Andy Clark has no clear principled reason for allowing cognition into the world but not any element of consciousness. In particular, certain cases of empathic-like connection pass beyond anything that ordinarily passes as empathy, with its requirement (as Zahavi correctly points out) that the self/other distinction be maintained. Instead of an alignment of experiences in each agent, one has a single, joint experience that is carried away afterwards by each agent., once again operating as individual experiencers. The phenomenologists’ claim to “direct” experience is on to something important – and correct in removing theory of mind from its pedestal – but it should dare to be much more bold.

REFERENCES

- [1] Frederick Adams and Kenneth Aizawa, *The Bounds of Cognition*, John Wiley and Sons, 2008.
- [2] Ron Chrisley and Joel Parthemore, ‘Synthetic phenomenology: Exploiting embodiment to specify the non-conceptual content of visual experience’, *Journal of Consciousness Studies*, **14**(7), 44–58, (2007).
- [3] Andy Clark, *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press, Oxford, 2008. Also available as ebook through Oxford University Press (<http://www.oxfordscholarship.com>).
- [4] Andy Clark, ‘Spreading the joy? why the machinery of consciousness is (probably) still in the head’, *Mind*, **118**(472), 963–993, (2009).
- [5] Andy Clark and David Chalmers, ‘The extended mind’, *Analysis*, **58**(1), 7–19, (January 1998).
- [6] M. Hogenboom. Psychopathic criminals have empathy switch. <http://www.bbc.com/news/science-environment-23431793>, July 2013. Last accessed 15 December 2016.
- [7] Joel Krueger, ‘Merleau-ponty on shared emotions and the joint ownership thesis’, *Continental Philosophy Review*, **46**, 509–531, (2013).
- [8] Victor Loughlin, ‘Sketch this: Extended mind and consciousness extension’, *Phenomenology and the Cognitive Sciences*, **12**(1), 41–50, (March 2013).
- [9] Henry Markram, Tania Rinaldi, and Kamila Markram, ‘The intense world syndrome: An alternative hypothesis for autism’, *Frontiers in Neuroscience*, (October 2007). <https://doi.org/10.3389/neuro.01.1.1.006.2007>.
- [10] Humberto R. Maturana and Francisco J. Varela, *The Tree of Knowledge: The Biological Roots of Human Understanding*, Shambhala, London, 1992.
- [11] Joel Parthemore, *Concepts Enacted: Confronting the Obstacles and Paradoxes Inherent in Pursuing a Scientific Understanding of the Building Blocks of Human Thought*, Ph.D. dissertation, University of Sussex, Falmer, Brighton, UK, March 2011. Available from

¹¹ Enactivist Tom Froese does, indeed, take up exactly this possibility in a draft chapter for the forthcoming *4E Cognition Handbook* from Oxford University Press. Unlike the position I have taken here, he argues that the question *can* be addressed empirically.

¹² ... A point I owe to Dan Egonsson.

- <http://www.sussex.ac.uk/Users/jep25/papers/thesis/thesis-a4.pdf>.
- [12] Joel Parthemore, 'Autism as philosophical insight: The enactive response to the tendency to pathologize', in *Proceedings of the AISB Symposium on Reconceptualizing Mental Illness, 3-5 April 2013*, pp. 12–16. Society for the Study of Artificial Intelligence and Simulation of Behaviour, (2013).
 - [13] Joel Parthemore, 'The unified conceptual space theory: An enactive theory of concepts', *Adaptive Behavior*, **21**, 168–177, (2013).
 - [14] Joel Parthemore, 'Beyond objectification: From robots as sex toys to a new theory of personhood', in *Proceedings of the AISB Symposium on Artificial Sexuality, 4 April 2016*. Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), (2016). <http://aisb.org.uk/publications/proceedings/aisb2016.zip>, Artificial Sexuality, paper 4.
 - [15] Joel Parthemore, 'Consciousness, semiosis, and the unbinding problem', *Language & Cognition*, (2017). In press. Corrected proofs available at <http://dx.doi.org/10.1016/j.langcom.2016.10.004>.
 - [16] Robert Rupert, *Cognitive Systems and the Extended Mind*, Oxford University Press, 2009.
 - [17] Dan Zahavi, 'Empathy, embodiment and interpersonal understanding: From lipps to schutz', *Inquiry*, **53**(3), 285–306, (2010).