

Empathy's blind spot

Jan Slaby

© Springer Science+Business Media Dordrecht 2014

Abstract The aim of this paper is to mount a philosophical challenge to the currently highly visible research and discourse on empathy. The notion of empathetic perspective-shifting—a conceptually demanding, high-level construal of empathy in humans that arguably captures the core meaning of the term—is criticized from the standpoint of a philosophy of normatively accountable agency. Empathy in this demanding sense fails to achieve a true understanding of the other and instead risks to impose the empathizer's self-constitutive agency upon the person empathized with. Attempts to 'simulate' human agency, or attempts to emulate its cognitive or emotional basis, will likely distort their target phenomena in profound ways. Thus, agency turns out to be empathy's blind spot. Elements of an alternative understanding of interpersonal relatedness are also discussed, focusing on aspects of 'interaction theory'. These might do some of the work that high-level constructs of empathy had been supposed to do without running into similar conceptual difficulties.

Keywords Empathy · Agency · Perspective-taking · Commitment · Interaction · Peter Goldie

Introduction

Steep indeed is the rise of empathy from what had been a tentative neologism in the translation of an academic work to a 'global force' that is assumed by some to be epoch-defining. Nothing less than *The Age of Empathy* (de Waal

2009) or *The Empathic Civilization* (Rifkin 2009) have recently been proclaimed. It would be worth a paper of its own to explore this astonishing career of the concept of empathy, track its historical trajectory and situate the empathy trend within broader cultural and intellectual developments. In the present paper, however, I will remain within a more narrow and exclusively philosophical register with the aim to elaborate upon a major structural obstacle that befalls at least one currently prominent conception of empathy. I will set the stage with a short overview of issues concerning the definition of central forms of empathy and of some key research approaches ("Setting the stage: What is empathy" section), before I develop a philosophical line of critique that concerns high-level cognitive empathy or what is sometimes called "empathic perspective-shifting". In this central part of the paper ("Why empathy fails: the usurpation of agency" section), I will undertake to show, by taking up and developing arguments initially brought forth by Moran and Goldie, that this demanding mental operation fails to get at what it is supposed to reach: at another person's experiential reality, at her individual perspective on the world. Instead, on the currently most influential understanding of empathy, the empathizer will likely only project and impose *her own* mental life, most notably her own agency, onto the other person. At the heart of the matter lies a deep-seated misconstrual of the structure of self-conscious human agency in general. Agency is empathy's blind spot. Thus, ultimately, a flawed philosophy of mind stands behind this highly problematic account of empathy. In the final part ("Alternatives: interaction and recognition" section), I will explore a theoretical alternative to the mainstream understanding of cognitive empathy, namely embodied *interaction theory* and related conceptions—approaches that have in its center the notion of a marked co-presence of

J. Slaby (✉)
Institute of Philosophy, Freie Universität Berlin,
Habelschwerdter Allee 30, 14195 Berlin, Germany
e-mail: jan.slaby@fu-berlin.de

individuals as emerging from joint engagements in projects or other forms of joint agency as opposed to inquisitive attempts of one person trying to ‘get at’ or ‘mirror’ another person’s mental states.

Setting the stage: What is empathy?

Low-level versus high-level empathy

In 1909, US psychologist Edward Titchener first employed the term “empathy” as a translation of the German term “Einfühlung” that figured in the work of German psychologist Theodor Lipps (see Titchener 1909). Interestingly, Lipps himself was the translator of David Hume into German, and used the term “Einfühlung” for Hume’s “sympathy”. Sympathy, in turn, quickly leads us back to Adam Smith, in whose work the term played an even more prominent role as in Hume’s. The German *Einfühlung* literally means “feeling into” another, a metaphorical construct that mobilizes various connotations and associations which surround the whole family of concepts used to refer to the different forms of interpersonal relatedness.

Given this conceptual lineage, it can be helpful to go all the way back to Smith for an initial characterization of something close enough to today’s empathy. It is of little relevance that Smith here speaks of “sympathy”—as the history of the term shows, what he drives at is what writers in the twentieth century predominantly class as “empathy”. Here is what Smith wrote:

By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feel something, which, though weaker in degree, is not altogether unlike them. (Smith 2002 [1759], p. 9).

This short passages displays a quite remarkable conceptual struggle already—forecasting difficulties that are still with us today. Smith seems undecided here between different conceptual options. What we can take from this is the central role the imagination plays in empathetic processes—the idea of an ‘imaginative placement’ *into* the other person’s situation, and this is obviously a complex process that requires the exercise of sophisticated cognitive capacities. But then it is already unclear what *exactly* this means. Smith, at any rate, seems to oscillate between imagining *oneself* as being in the other’s situation and between the more ambitious idea of *becoming the other person*—i.e. something like imagining *being the other*.

As one can easily glean from wealth of present-day work on empathy, not everything has been cleared up and

made precise since the times of Hume and Smith.¹ It is striking how many different conceptual and theoretical approaches to empathy can be found in the literature.² A rough ordering might be imposed on the material by distinguishing between *lower-level* theories that are usually quite inclusive as to what they count as empathy, and *higher-level* theories that place much stricter—usually cognitive—demands on candidate processes.

Primatologist Frans de Waal, author of many scientific and some popular works on empathy, is an exemplary proponent of a very inclusive approach to empathy that starts out with low-level mechanisms. In his ‘Russian doll model’ of empathy—so called because of the nested structure that has evolutionary older mechanisms retained in more sophisticated newer ones—the spectrum of empathic processes ranges from quite basic ones of state matching such as emotional contagion, affect attunement, and bodily resonance via processes such as sympathetic concern all the way up to full blown cognitively mediated perspective-taking (cf. de Waal 2009, p. 208/9). Strategy-wise, de Waal follows in the footsteps of Darwin (1981 [1871]). In notable contrast to his coarse-grained popularizer Thomas Huxley who advocated a grim view of human nature, Darwin was a firm believer in deeply rooted moral instincts in humans, developing out of capacities to care for those in one’s group and to cooperate with peers that Darwin assumed operative throughout the higher ranks of the animal kingdom. In about the same manner, a key line of de Waal’s work is to search for altruistic, cooperative, and reconciliatory behavior in primates. The assumption is that these are then nested within—and still functionally contributing to—evolutionary younger, cognitively sophisticated mechanisms. While the model has some appeal in its reckoning with a structural continuum of a broad range of other-related traits, its enormous inclusivity does little to resolve the problem as to what actually should be the central meaning of the term ‘empathy’ in current research contexts. If traits as distinct as basic affective state matching and as sophisticated as cognitive perspective-taking are lumped together under the same umbrella, there is rather little specificity in what counts as empathy—items collected under the term’s scope would range from the most automatic to the most deliberate of mental processes. The concept would thus seem like a mere container concept subsuming a heterogeneous lot of mental processes.

¹ I should briefly mention in what way Hume’s view on “sympathy” differed from that of Smith. Chiefly, it can be said that whereas Smith pointed towards a more demanding mediation by the faculty of imagination, Hume rather foregrounded automatic processes such as the unreflective mirroring of another person’s emotions (cf. Coplan and Goldie 2011, p. x–xii).

² Good overviews and attempts at systematization are provided by Batson (2009), Coplan and Goldie (2011) and Stueber (2013).

Empathy as perspective-shifting

A much-needed precisification of the prevailing understanding of empathy is courtesy of some conceptual groundwork by philosophers Coplan and Goldie. Coplan in particular presents a convincing case for why we should—pace de Waal—distinguish higher-level, conceptually demanding cases of perspective shifting from a variety of lower-level phenomena. Without this clarification, we would be at a loss to understand and assess the various empirical findings on issues such as mirror neurons, shared representations, emotional contagion, mental simulation, and so on. Terminological ambiguity and conflation of quite distinct types of process have hampered the prospects for assessing the progress in empathy research (cf. Coplan 2011, p. 6). Coplan then specifies her own conceptualization of empathy as follows:

[E]mpathy is a complex imaginative process in which an observer simulates another person's situated psychological states while maintaining clear self-other differentiation. To say that empathy is 'complex' is to say that it is simultaneously a cognitive and affective process. To say that empathy is 'imaginative' is to say that it involves the representation of a target's states that are activated by, but not directly accessible through, the observer's perception. And to say that empathy is a 'simulation' is to say that the observer replicates or reconstructs the target's experiences. (Coplan 2011, p. 6/7)

Thus, Coplan's proposal singles out three component processes that jointly make up fully-fledged empathy: affective matching, other-oriented perspective taking, and self-/other- differentiation. As we will see below, all these components pose difficulties. What is crucial at this point is that there are good reasons for reserving the term "empathy" for exactly this kind of complex, demanding process. Not doing so would threaten to blur our reasonable commonsense understanding of empathy as not just *any* affectionate process directed at or involving another's state of mind, but as a precise way of 'feeling-into' another's perspective and thus coming to know the other's predicament from the inside, without losing one's grip on the distinction between ourselves and the other. As Coplan persuasively shows, even a process as demanding as *self-oriented* perspective taking—imagining *oneself* in another person's shoes—wouldn't do, because of the usually profound, deep-seated differences between individuals. What *oneself* would feel in a given situation is for the most part not a very good measuring stick of what another person would feel—outside of a limited range of mostly trivial cases. The problem is made more severe by the default tendency to assume greater similarities between people

than there in fact are—egocentric bias is a pervasive psychological trait in humans (cf. Savitsky 2007). Accordingly, Coplan concludes:

To summarize, personal distress, false consensus effects, and general misunderstandings of the other are all associated with self-oriented perspective-taking. When we imagine ourselves in another person's situation, it frequently results in inaccurate predictions and failed simulations of the other's thoughts, feelings, and desires. It also makes us more likely to become emotionally over-aroused and, consequently, focused solely on our own experiences. (Coplan 2011, p. 14).

The only viable alternative conceptualization is other-oriented perspective-shifting. Goldie christens it 'empathetic perspective-shifting' and specifies it as follows: "Consciously and intentionally shifting your perspective in order to imagine *being* the other person, thereby sharing in his or her thoughts, feelings, decisions and other aspects of their psychology." (Goldie 2011, p. 303). Accessing another's mind *from the inside*—and thus only producing the same mental states in oneself as one assumes the other person to have, but shifting imaginatively into the other's predicament while maintaining a clear-cut self/other differentiation.³ Only then, or so the expectation goes, might one succeed to feel what the other feels not from one's own perspective but from the other's. Only then will one 'get at' what one wants to get at in one's earnest attempts at understanding another person. But now a crucial question arises: Can this quite demanding mental maneuver really be *successfully* executed?

Why empathy fails: the usurpation of agency

How might one possibly do this: imagining *being* the other—without inevitably projecting and thus imposing what is in fact our own 'being' onto the other person? Can we really *successfully, genuinely* imagine being another person?

I am highly pessimistic here, and with this I very much side with Goldie (2011), who sadly didn't live to develop his profound critique of empathy any further. Goldie sets

³ This is how Coplan justifies the additional condition of a clear self-other-differentiation: "Without clear self-other differentiation, we are almost certain to fail in our attempts to empathize. We either lose our sense of self and become enmeshed [with the other's experience] or, more often, we let our imaginative process become contaminated by our self-perspective and thus end up engaged in a simulation that fails to replicate the experience of the other. Self-other differentiation allows for the optimal level of distance (...). We are neither fused nor detached." (Coplan 2011, 17).

out to demonstrate that, outside of very primitive cases, empathetic perspective-shifting simply fails. It is just not possible to do it in the way needed to realize the demanding construal of empathy. But, and that is the good news, it is also not required for civilized interpersonal conduct. The fact that empathetic perspective shifting doesn't work is not tragic. Rather, its failure is instructive, because in analyzing it we learn a philosophical lesson about what it means to be a full-blooded agent, about what it means to possess a practical point of view. Understanding this failure provides us with a more adequate understanding of the mind and of personhood, and thus is in the end also informative for a better way to conceive of beneficial and praiseworthy ways of interpersonal interaction that actually *do* work. As I hope to make plausible presently, an adequate understanding of agency's involvement in the mental is the linchpin around which several core controversies in the philosophy of mind currently revolve. It is crucial to make this widespread combination of an underestimation and a misconstrual of agency explicit and point out its problematic consequences.

Outside the base case

This is what Goldie calls the 'base case'—the standard shape of an interaction scenario that is assumed as a template for the explication of empathetic perspective shifting: A person 'A' attempts to empathize with another person 'B' in a situation where there are

- (i) no relevant differences in the psychological dispositions of A and of B
- (ii) no relevant nonrational influences on B's psychological make-up or decision-making process;
- (iii) no significant confusions in B's psychological make-up; and
- (iv) no psychological conflicts (where B has to make a choice where it is not clear to B which alternative is to be preferred). (see Goldie 2011, p. 308)

The point is, for this kind of case, empathic perspective shifting might indeed be supposed to work—but that's just because about everything interesting has been omitted here! Base case scenarios are almost completely uninteresting when it comes to non-trivial personal affairs. These are cases so simple and obvious that even largely uninformed and insensitive persons could arrive at correct predictions as to what is on B's mind and what B is going to do next. Crucially, in base case scenarios, predictions can be made *regardless of what person it is* whose mental states are to be predicted (cf. Goldie 2011, 309). This is not nothing, obviously, and sometimes one indeed empathizes successfully on these grounds, but cases like these are a far

cry from what the current wave of empathy research is chiefly aiming at.

The problem is that when it really matters to understand another, cases are usually not like the base case. Conditions (i) to (iv) won't be fulfilled together; in fact, all four of them are likely to be violated in most meaningful real-life encounters. Attempts to empathize have to deal with relevant differences in disposition, ability and character; the prospective empathizer will have to countenance nonrational influences on the target person's thinking and deciding; and likewise, we will regularly have to assume a certain level of confusion (or of uncertainty and indeterminacy) in any given person about her own attitudes and experiences. Not least, quite often there will be conflicts calling for nontrivial resolution, such as when important, nonobvious choices have to be made. I will now briefly review Goldie's case by case treatment of the four points that distinguish most real life scenarios from the base case.

First, consider the way a person's characteristics, such as being friendly, shy, or timid, partly shape what is on her mind at a given point while not figuring among the conscious contents of her mental states. Obviously, character traits have a massive impact on what a person is thinking and feeling at a given point in time—but this impact is for the most part tacit and in the background of awareness, it is not something that a person is consciously focused on, let alone something she freely and willfully deploys. As Ryle aptly said: "The vain man never feels vain" (Ryle 2000 [1949], p. 87). So given the not unlikely case that person A possesses markedly different character traits than person B, how will she be able to provide for this difference in her simulation attempt? Here is Goldie's description of the predicament:

A cannot, as part of a consciously willed project, keep B's characterization in the non-conscious background (...). A will be obliged, in trying to shift to B's perspective, to treat B's characterization through the theoretical or empirical stance, as one typically does when considering the role of character in explaining or predicting other people's decisions, actions. (Goldie 2011, p. 309)

Thus, we arrive at the first sense in which the agency of person B is likely to be 'usurped' by the empathizer: Background dispositions, traits and abilities of B, as far as these are known, have to be objectified and introduced consciously and deliberately into the process of imaginative perspective shifting. Short of some kind of professional method acting, this will inevitably ruin the attempt to imagine *being* the other person, as person A will have to artificially bring to bear alien character traits in order to frame person B's states of mind in the right way.

A similar problem is raised by the second usual difference between individuals, viz. the routine amount confusion and irrationality that besets most natural-born human minds. For instance, how to mentally reproduce another person's *unconscious* mood-related irrationality? (cf. Goldie 2011, p. 313) Again, the most likely and most practicable option is for A to *artificially objectify* the irrationality-inducing state of feeling. For example, A could try to introduce B's anxious and timid background feelings into her simulation attempt by actively bringing herself in a similarly anxious state of mind—however, if successful, this would likely yield a foreground feeling of fear but not a pervasive background anxiety, such as an existential feeling of fearfulness (Ratcliffe 2008).⁴ Again, nothing short of method acting would let the empathizer reach a mental state of the kind that likely shapes B's experiential background.

There is a common pattern to all these problems, and it comes into clearer focus when we consider points (iii) and (iv) on Goldie's list. Mental states are not like 'inner objects' that appear on some kind of inner stage or mental 'video screen'. The issues so far discussed have shown that the background frame of mind—character traits and background moods pre-consciously affecting a person's mental life—cannot be objectified and introduced *alongside* conscious foreground mental states. A person's conscious perspective has a background/foreground or projection/baseline structure that is virtually impossible to emulate by another. What the subsequent points will show is that objectification *never* works for mental states, regardless of foreground or background, fully conscious or hidden in the background of awareness. Mental processes are notoriously not the 'objects' of some inner perception—rather, they are that *through which* a person directs herself to the world, dynamically, actively, from her own often unique perspective. The basic situation is thus one of active, prospective engagement with the world: a future-directed *positioning* towards what goes on. Concretely, this means that a mental agent for the most part has an active say in determining her

⁴ The central idea behind Ratcliffe's (2008) proposal is that of a fundamental *felt relatedness* of self and world. As "ways of finding oneself in the world", felt existential orientations establish the most basic self/world mutuality in experience—long before conscious reflection sets in and even at a point "before" it makes sense to separate at all a subject of experience from the world it confronts. Examples for existential feelings are feelings of connectedness to the world, feelings of familiarity and security, feelings of belonging to a group or to other people in general, feelings of being in control and feeling capable of this or that action, but also a quite general sense of the "being" of worldly entities, of oneself, and of others as fellow persons. The key for present purposes is that existential feelings are usually so deeply engrained in a person's overall perspective on the world that one is at a loss as regards their simulation or deliberate "deployment" (see Slaby and Stephan 2008 for a more detailed discussion of Ratcliffe's proposal).

own subsequent mental processes. As beings capable of self-interpretation (Taylor 1985), we have a say about what it is that we think, feel, or want. This becomes especially clear in the case of confusion, as Goldie explains:

The ability to reflect on our confusion, and decide what we think or feel, has at its heart the full-blooded notion of agency in relation to our own minds. Thoughts are thought, feelings felt, decisions and choices made, *by particular agents*, and the identity of the agent in this full-blooded sense can make a difference to what is thought, felt, decided on, or chosen. It is not as though all thoughts, feelings, decisions, and choices can be 'processed' by *any* agent, impersonally... (Goldie 2011, p. 315)

Thus, to be a conscious and self-conscious agent—in other words, a person—is not like being a container filled with a peculiar sort of entities or states, viz. mental states, as some class of determinate, observable objects. Being an agent means having a say in specifying, in ultimately deciding and *committing* to what one will have on one's mind. In the absence of a capacity to actively get a hold of what unfolds in one's consciousness, to be able to guide, however partially and tentatively, some of one's own mental stream, we would be at a loss to identify a mind as we know it. As an example, take the case of belief—a mental state kind that is absolutely basic for human mentality: Asked what you believe about some matter, you surely won't respond by looking *inside yourself*, in order to somehow grasp what it is that you believe about the matter in question. Rather, upon being asked, you begin to actively think about the world, orient yourself toward the issue under discussion, and make up your mind about it. You deliberate, you judge and weigh options and thereby you shape, *actively*, your belief. And part and parcel with this comes a readiness to defend your belief against challenges or revise it in light of new evidence—all these are *active* capacities that are inextricable from your capacity to form and entertain beliefs.⁵ Or take the different but equally relevant mental state type "emotion": Imagine a complicated social situation in which you might not initially know what you are feeling—is this really a case of having to introspect more carefully in order to *find out* what that 'true feeling' is that you're having? Or is even this more often rather a *practical* question where it is partly up to you to determine what your feeling is, what is appropriate, what the situation demands? As Charles Taylor

⁵ Note that this does not entail that belief is active in the sense of voluntarism. I cannot believe 'whatever I want to believe' exactly because I am actively committed, qua my capacity to belief, to only believe what I deem true. One might say that as a believer, I am *actively bound* by truth. The activity in question concerns the capacities to adjust my attitudes in accordance with what I discover to be the case (see also Moran 2001, 51–55).

has repeatedly shown: Even our confusions and self-misunderstandings enter into our mental states, and come to shape their content and direction (Taylor 1985).⁶

The same point comes again strikingly to the fore in the case of manifest conflict (Goldie's fourth point), where a person is pulled in different directions in a situation of choice, and thus called upon to decide between competing options. Surely, those decisions have to be radically first-personal in that one cannot just neutrally ponder one's options, calculate their pros and cons, look their respective fixed values up in a table of preferences, as it were—and then just decide according to the result of that calculation. Instead, one inevitably will have to *bestow*, first-personally, weights upon one's options. The agent will have to adopt a stance, undertake a commitment, emphatically bring herself to bear in resolving the issue. While rational, and thus partly the outcome of deliberation, personal choice has to concretize value in one's own given case. Consider what Paul Ricœur writes on this matter:

It is of the essence of value not to appear except as a possible motive of decision. I testify to a value only as its champion. (...) I see only those values which I am willing to serve. (Ricœur 1966 [1950], p. 75)

This element of commitment—genuine choice in terms of first-personal involvement—is what escapes attempts of deliberately simulating another's mental states: “in empathetic perspective-shifting, where A thinks B's thoughts, and then in imagination decides what is the right thing for B to think or to do, A *usurps* B's agency, replacing it with her own” (Goldie 2011, p. 315). It is agency that is needed in making a choice—and it is agency in the rich sense of bringing oneself to bear on a given matter. And this agency, in a situation of empathetic perspective shifting,

⁶ One reviewer asked who the “you” is that is called upon to determine what it is that is felt in a given moment. Given the fundamental role of agency for human mentality, one might be tempted to assume, in a Kantian spirit, that this cannot be the fully constituted empirical subject and instead has to be some kind of transcendental ego that is there at the very beginning, an instance without qualities except the bare capacity to act, and then self-constitute successively in a series of empirical acts. In line with Heidegger and Sartre, I think this would be a grave mistake: There is only one agentive instance, and this is the empirical person with all her contingent characteristics already in place. What often happens, however, is that we do not stand up to the challenge of acting (deciding, resolving conflicts) in a given situation, let alone in a situation of emotional turmoil, but instead let things just run their course. This inauthentic, routine everyday mode of *mauvaise foi* can thus mask the fact that our agency is constantly “there” as long as we are conscious, making us constantly able to take charge and intervene in a current situation, including so as to resolve upon an emotional response adequate to it. Thus, obviously, and as will become clearer below, the account of agency given here is very much in the tradition of existential phenomenology. On the specific case of “activity” in regard to feeling and emotion, see Sartre (1994 [1939]).

can only ever be the agency of the empathizer, not the agency of the person empathized with.

So this is what Goldie means by the ‘full-blooded notion’ of agency. A full-blooded agent is not somehow ‘separate’ from her own mental states as if these states were inner objects swimming in some kind of impersonal mental stream. Instead, the agent actively brings about what she has on her mind, at least to a significant extent. An agent is the very ‘instance’ capable and called-upon to *commit*, to adopt stances towards the world and be ready and willing to live up to them (beliefs, decisions, attitudes)—the mental cannot be understood in abstraction from this domain of commitment. The common idea of first person authority about one's mental states is not to be understood as an epistemically privileged position, as a matter of just ‘knowing’ one's own, as it were, pre-existing mind, but it is crucially a matter of first-person *authorship* with regard to our attitudes and stances (Moran 2001). How can this be supposed to be simulated by someone else without this other person inevitably *imposing* her own interpretations, decisions, commitments, or would-be commitments? The core of what it means to be a genuine agent—the instance that, however confused, fleeting and instable, has a say in determining what is thought or felt—cannot be empathetically simulated without an alien imposition, without an artificial act of objectifying or imposing the empathizer's own agency. Trying this in earnest would be a move that comes close to patronizing the other because one inevitable will take what is in fact *one's own* agency (or would-be agency) for the agency in which the other person's mental states are anchored. As agents, we are in an important sense *irreplaceable*: fully-fledged agency is in each case essentially someone's, there is an ineliminable moment of authentic ownership, or *Je-meinigkeit*, to use Heidegger's term (Heidegger 1927, p. 41).

We see here that this line of thought espoused by Goldie reaches back, via Moran's insightful account of self-knowledge (Moran 2001), at least to the existentialist tradition, Sartre and Heidegger. Being an agent means indeed this, *being it yourself*, nobody can take over your agency, and decide and live in your stead. As helpfully put by Schear, this has the consequence that being a human subject means “being able to fail at being a subject” (Schear 2009, p. 105)—human subjectivity is an existential task, something we have to actively undertake and keep undertaking as long as we exist as self-conscious subjects.

Thus, and importantly, Goldie's critique has a much wider scope than just the debate surrounding empathy and interpersonal relatedness more generally. His critique of accessing *other* minds through empathy mirrors in key respects a critique of a standard construal of accessing one's *own* mind through introspection: Our mental states are not like ‘inner objects’ that we can either perceive accurately or misperceive—as has been claimed by some recent psychological defenders of introspection as a scientific method (Hurlburt

2009, for a critique, see Schwitzgebel 2008, see also Hurlburt and Schwitzgebel 2007). As self-interpreting, self-constituting agents, we have a notable, if obviously limited, first-personal *authority* in deliberating about and determining *what it is* that we think, feel, or want. Thereby, we contribute to the 'making of' our mental states by deliberating on our situation and by adopting and sustaining attitudes towards what is going on. Attentively making up one's mind and committing to an attitude—a belief, say—are *agentive* capacities, active exercises of abilities. These have limitations, obviously—they have to be exercised on the basis of materials that are pre-given and they often have to be executed and subsequently upheld against various resistances, requiring resolve and effort. But ultimately, the key point is that the mental is in the register of what we do—and not something that merely wells-up, passively, within us. This basic agentive core of a person's perspective on the world needs to be protected against attempts to ignore it or explain it away. Accounts that give pride of place to the automatic, the involuntary, the passive aspects of the mental can thus ever only be a part—and, importantly, only a non-autonomous part!—but never the whole of the story about the human mind.

It should be clear that these considerations will not doom *every* kind of interpersonal understanding, let alone basic unreflective processes of attunement and resonance, but what should now be evident is that the right stance towards another person is not one of trying to reproduce, within one's own mind, the mental states of the other.

Alternatives: interaction and recognition

Recent phenomenological work on what has become known as *interaction theory* has the potential to make progress in this regard, offering a possible alternative to empathetic perspective-shifting.⁷ It is no accident that interaction theory thoroughly breaks with simulation accounts of empathy (see, e.g., Gallagher 2008, 2012). This family of alternative approaches, instances of which having been developed by authors like Gallagher, Krueger, Fuchs, Ratcliffe and others, does not conceive of the understanding of another's mind as a case of one person simulating or mirroring an other's mentality, but instead focuses on mutual interactive engagement and on marked forms of co-presence or 'being-with' one another. A key ingredient of interaction theory is the idea that mental states are for the most part not 'hidden' in the interior

⁷ I will not here discuss Zahavi's interesting and original phenomenological critique of empathy-as-simulation approaches. To Zahavi, the cognitively demanding form of empathy discussed here rests on a primordial form of intersubjectivity which is regularly missed by authors in the simulation tradition, so that they mistakenly think that our access to another person's mentality can only be established through such high-level simulative activities; see, e.g., Zahavi (2001).

of another person, but instead something that is in plain view, and might thus be experienced directly and without difficulties. Gallagher in particular has argued insistently against the need to invoke mental simulation or other demanding forms of perspective-shifting and in favor of assuming access to another's mind to be a matter of such perceptual take-up (e.g., Gallagher 2008). Crucially, for present purposes, the context in which such perception (or experience more broadly) of mental states takes place is not detached, stand-off-like relationships but situations of ongoing interaction in practical contexts. Here—amidst the joint engagement in a project or the shared involvement in social event—is usually not much of a riddle as to what another person is currently thinking, intending or feeling. There is no need to assume an intricate inquisitive endeavor of one person getting 'at' or even 'into' another mind. Instead, forms of *joint* agency, and *joint* active world-orientations make up the background against with a smooth interactive relatedness unfolds. With this, interaction theorists appreciate what has been shown in the previous section: that the only way to meaningfully engage with another person's mentality without imposition is by engaging with her on the level of action—establishing a kind of *co-engagement*, as it were, for example by jointly striving toward some goal or by jointly enacting a project. "Participatory sense-making" has been one helpful concept developed for this purpose (de Jaegher and di Paolo 2007), another is the phenomenological concept of a "we-space" (Krueger 2011)—a realm of co-presence, of lived mutuality, bodily enacted between interacting individuals, a transformation seamlessly imposed upon one's surroundings once joint agency is established. Within the many varieties of joint active engagements, the partners of interaction come to achieve a kind of felt coalescence that brings forth exactly the kind of 'union of minds' that advocates of empathy strive for—not by way of mental simulation but by first building up a joint perspective on the world that both partners may identify with and adopt as their own. Within this shared perspective, experiential responses to intentions, desires, thoughts, feelings and, above actions of the partner are enacted, albeit not in the form of a succession of discrete mental states but in the manner of a seamless relating inextricable from the ongoing unfolding of the joint activity. Importantly, what is operative in these situations is an active, constructive and forward-looking orientation. It is this shared 'looking ahead' towards goals, a shared anticipation of likely developments and events which constitutes the joint perspective and lets an experiential "we space" open up.⁸

⁸ What must be added to the interactionist perspective is an background understanding of social situations and interaction contexts in terms of narrative, and accordingly a narrative competency, i.e. the ability to relate events and happenings within meaningful common contexts. For more on narrative as an element of interaction theory, see Gallagher (2012) and Gallagher and Hutto (2008).

What we do not need here is miraculous ‘bridge’ between two initially separated and closed-off inner realms.⁹ A helpful concept that has been suggested in this context is that of ‘caring-with’, which builds upon Heidegger’s notion of care (‘Sorge’ in German) as the basic mode of a person’s ‘concernful’ world-orientation, but going further than Heidegger by showing that this very basic form of active, purposeful striving might from the outset be shared between two or more individuals (for an elaboration of this idea, see Sánchez Guerrero 2011).¹⁰

What this above all shows is that the quasi-inquisitive attempt to fully ‘get at’ another person’s perspective on the world is superfluous. Nothing is lost when one replaces this by a stance of *acknowledging*, of *recognizing* the other, both in her (partial) agentive autonomy and in her exposedness as a vulnerable, needy being (cf. Butler 2001). To be sure, with this thought we already move beyond the rather narrow discussion of nature of understanding another’s perspective on the world, but it is important to broaden the scope of the account in this way. Recognizing the other, as part of ongoing joint active engagement with the world, is what opens up a truly human perspective upon one another. Adopting this stance towards another person, we come to *let her be* in what in many cases may well remain an inevitable alterity. We thus drop the assumption that the goal of interpersonal relatedness would inevitably have to be an encompassing understanding of the other person, a kind of ‘total transparency’. ‘Letting be’, accepting, acknowledging the other—this might actually be what Descartes, of all people, had in mind when he spoke of *générosité*: The truly ‘generous’ stance of transcending one’s own partial, enclosed perspective in order to open oneself up and let oneself be ‘conditioned’ by the other (see Ricœur 1966 [1950], p. 62). Regardless of Descartes’ surprising involvement in articulating this idea, what has definitely been left behind here is the conception of closed-off inner subjective mental realms. Instead, the mental comes in view as an active openness to the world and to others—an openness that ‘awaits’ its being conditioned, challenged, saddled with contents to which it will then adopt certain attitudes on the basis of which it will further go on to be this ‘lived’ open dimension. In

comparison with this dimension, construals of ‘inwardness’—i.e., an exclusive space of private ‘inner states’—are most likely a rather late-coming, culturally contingent achievement, an objectification of the original spontaneity of the human mind.

To come full circle with this as a model for an open, responsible togetherness in mutual recognition, we would have to ground the interactionist approach within something like Judith Butler’s perspective on the human subject’s irredeemable relationality and mutual dependency. This provides a depth dimension which brings the still abstract image of an agentive openness to the world and to others into accord with its inevitable existential foundation as epitomized in our needful exposure towards one another as finite and vulnerable beings:

What might it mean to make an ethic from the region of the unwilling? It might mean that one does not foreclose upon that primary exposure to the Other, that one does not try to transform the unwilling into the willed, but to take the very unbearable of exposure as the sign, the reminder, of a common vulnerability, a common physicality, a common risk. [...] [Violence] delineates a physical vulnerability from which we cannot slip away, which we cannot finally resolve in the name of the subject, but which can provide a way to understand the way in which all of us are already not precisely bounded, not precisely separate, but in our skins, given over, in each other’s hands, at each other’s mercy. This is a situation we do not choose; it forms the horizon of choice, and it is that which grounds our responsibility. In this sense, we are not responsible for it, but it is that for which we are nevertheless responsible. (Butler 2001, 39)

Besides much else, Butler here restores *passivity* to its proper place in human existence and thereby limits the power, though not the reach, of agency. Our agency is at any time bound and conditioned by the inexorable givenness and inevitability, in each concrete instance of human existence, of an essentially needy organic base in its dependency and vulnerability. In this sense, one can say that we are at the mercy of one another, and with this we have anchored an existential mutuality on a very ‘deep’ level of thought on our human condition.

Again, these last considerations are a long way removed from the specialist debates about the mechanisms of high-level empathy, and I cannot here go into more detail about the complex alternative narrative of existential mutuality, whose outlines I have barely sketched. What should have become clear even from these sketchy fragments is that there is no need to let a narrowly-focused ‘mental simulation’ literature monopolize the difficult and important topic of interpersonal relatedness.

⁹ In elaborating this further one would have to include lower-level mechanisms of attunement and resonance into the theory, so as to find room for a pre-cognitive relationality—probably established and sustained on the level of affect and affect attunement (see Stern 1985)—that sets the stage for more sophisticated forms of interactive engagement. Proponents of interaction theory attempt to supply such a foundation (Froese and Fuchs 2012; Krueger 2009).

¹⁰ I keep the discussion of interaction theory deliberately brief as it is not in the center of the present account. Obviously, much more complexity has to be added to these outlines in order to arrive at a viable theoretical account. See Ratcliffe (forthcoming) for a discussion of some aspects of interaction theory from a perspective similar to the one adopted here.

Conclusion

I have argued that human agency is empathy's blind spot. According to a powerful strand of current philosophical theorizing on the nature of interpersonal relatedness, empathy should be understood as a demanding form of cognitive or imaginative perspective-shifting. However, when this account is adequate then we must concede that empathy is either largely useless (succeeding only in basic, uninteresting cases) or simply impossible—in the vast majority of cases of interaction that are not trivial because of differing character traits, deep-seated irrationality, or because non-trivial choice or conflict are in play. These complex cases illustrate that a person's mind—her experiential and practical perspective on the world—is inextricable from self-conscious and self-constitutive agency. The human mind is not a passive container filled with an array of 'mental items' that might then somehow be 'mirrored', 'simulated' or otherwise 'represented' by another. Rather, the human mind is essentially active—a matter of deliberating, choosing, of actively committing, of adopting and sustaining attitudes, or in short: a matter of actively 'living' one's life. What we are as persons is not just 'handed over' to us as a fixed and self-same identity, but an essentially first-personal *agentive* process we have to lead from moment to moment in the form of an active, responsible world-orientation. Only accounts that initially reify the mind's spontaneity into an array of fixated, ready-made 'inner states' might subsequently advance an understanding of empathy as perspective-shifting. By invoking the active character of our mental lives, the account given in the present paper points beyond expert discussions on the nature of empathy and towards other domains and debates in which the nature of the mental is at stake. Likewise, appreciating the indispensability of agency at the root of the mental clears the path towards a more adequate construal of the way in which persons are genuinely 'in touch' with one another. In this direction—interaction theory, accounts of joint agency and of mutual recognition—lie important further tasks for a philosophy of mind that aspires to move for good beyond all forms of solipsism, subjectivism, and individualism.

References

- Batson, C.D. 2009. These things called empathy: Eight related but distinct phenomena. In *The social neuroscience of empathy*, ed. J. Decety, and W. Ickes, 3–15. Cambridge: The MIT Press.
- Butler, J. 2001. Giving an account of oneself. *Diacritics* 31(4): 22–40.
- Coplan, A. 2011. Understanding empathy. In *Empathy: Philosophical and psychological perspectives*, ed. A. Coplan, and P. Goldie, 3–18. Oxford: Oxford University Press.
- Coplan, A., and P. Goldie (eds.). 2011. *Empathy: Philosophical and psychological perspectives*. Oxford: Oxford University Press.
- Darwin, C. (1981 [1871]). *The descent of man, and selection in relation to sex*. Princeton: Princeton University Press.
- De Jaeger, H., and E. Di Paolo. 2007. Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences* 6(4): 485–507.
- De Waal, F. 2009. *The age of empathy: Nature's lessons for a kinder society*. New York: Harmony.
- Froese, T., and T. Fuchs. 2012. The extended body: A case study in the phenomenology of social interaction. *Phenomenology and the Cognitive Sciences* 11: 205–235.
- Gallagher, S., and D. Hutto. 2008. Primary interaction and narrative practice. In *The shared mind: Perspectives on intersubjectivity*, ed. J. Zlatev, T.P. Racine, C. Sinha, and E. Itkonen, 17–38. Amsterdam: John Benjamins.
- Gallagher, S. 2008. Direct perception in the intersubjective context. *Consciousness and Cognition* 17: 535–543.
- Gallagher, S. 2012. Empathy, simulation and narrative. *Science in Context* 25: 355–381.
- Goldie, P. 2011. Anti-empathy. In *Empathy: Philosophical and psychological perspectives*, ed. A. Coplan, and P. Goldie, 302–317. Oxford: Oxford University Press.
- Heidegger, M. 1927. *Sein und zeit*. Tübingen: Max Niemeyer.
- Hurlburt, R. 2009. Iteratively apprehending pristine experience. *Journal of Consciousness Studies* 16(10–12): 156–188.
- Hurlburt, R., and E. Schwitzgebel. 2007. *Describing inner experience: Proponent meets sceptic*. Cambridge, MA: The MIT Press.
- Krueger, J. 2011. Extended cognition and the space of social interaction. *Consciousness and Cognition* 20(3): 643–657.
- Krueger, J. 2009. Empathy and the extended mind. *Zygon: Journal of Religion and Science* 44(3): 675–698.
- Moran, R. 2001. *Authority and estrangement: An essay on self-knowledge*. Princeton, UP: Princeton.
- Ratcliffe, M. 2008. *Phenomenology, psychiatry, and the sense of reality. Feelings of being*. Oxford: Oxford University Press.
- Ratcliffe, M. (forthcoming). The phenomenology of depression and the nature of empathy. *Medicine, Health Care and Philosophy*.
- Ricœur, P. (1966 [1950]). *Freedom and nature: The voluntary and the involuntary*. (trans: Kohák, E.V.). Evanston: Northwestern University Press.
- Rifkin, J. 2009. *The empathic civilization: The race to global consciousness in a world in crisis*. London/New York: Tarcher/Penguin.
- Ryle, G. (2000 [1949]). *The concept of mind*. Chicago: Chicago University Press.
- Sánchez Guerrero, A. (2011). Gemeinsamkeitsgefühle und Mitsorge. In *Affektive Intentionalität. Beiträge zur welterschließenden Funktion der Gefühle*, eds. J. Slaby, A. Stephan, S. Walter, and H. Walter, 252–282. Paderborn: Mentis.
- Sartre, J.-P. (1994 [1939]). *Sketch for a theory of the emotions*. London: Routledge.
- Savitsky, K. 2007. Egocentric bias. In *Encyclopedia of social psychology*, ed. R. Baumeister, and K. Vohs, 279. Thousand Oaks: SAGE.
- Schear, J.K. 2009. Experience and self-consciousness. *Philosophical Studies* 144: 95–105.
- Schwitzgebel, E. 2008. The unreliability of naïve introspection. *Philosophical Review* 117(2): 245–273.
- Slaby, J., and A. Stephan. 2008. Affective intentionality and self-consciousness. *Consciousness and Cognition* 17: 506–513.
- Smith, A. (2002 [1759]). *The Theory of Moral Sentiments*. Cambridge: Cambridge University Press.
- Stern, D.N. 1985. *The interpersonal world of the infant. A view from psychoanalysis and developmental psychology*. New York: Basic Books.
- Stueber, K. (2013). Empathy. In *The Stanford Encyclopedia of Philosophy* (Spring 2013 Edition), ed. Edward N. Zalta,

forthcoming. <http://plato.stanford.edu/archives/spr2013/entries/empathy/>.

Taylor, C. (1985). Self-interpreting animals. In *Human agency and language*. Philosophical papers, Vol. 45–76. Cambridge: Cambridge University Press.

Titchener, E.B. 1909. *Lectures on the experimental psychology of the thought-processes*. New York: Macmillan.

Zahavi, D. 2001. Beyond empathy: Phenomenological approaches to intersubjectivity. *Journal of Consciousness Studies* 8(5-7): 151–167.