Edith Steins's Philosophy: Implications for a Direct Functional Model of Empathy

A Filosofia de Edith Stein: Implicações para um Modelo Funcional Direto de Empatia

Eileen Pfeiffer Flores[1], Bianca da Nóbrega Rogoski[1]

Abstract: We define empathy, broadly, as understanding another person's experience. We begin by describing Edith Stein's phenomenological analysis of empathy. We argue for the continuing relevance of her critique of theories of empathy that presuppose the need for inferences from the "internal" to the "external" and show how mainstream psychological descriptions of empathy based on such assumptions have led to conceptual confusion and, ultimately, to deviance from the phenomenon of interest. We tease out the implicit account of empathy in Skinner's hypotheses about how we learn to describe private events. We argue that this account is characterized by a lingering inner-outer dualism that leads to incoherence when taken to its ultimate consequences. We propose an alternative conceptualization of empathy, the Direct Functional Model, with three principles: (1) primary givenness of experience, (2) priority of the whole, and (3) interaction. We argue that the model avoids the pitfalls of internal-external dualism and offers an account consistent with the philosophy of Radical Behaviorism.

Keywords: empathy, Edith Stein, radical behaviorism, theory of mind, perspective taking.

Resumo: Definimos empatia, de forma ampla, como a compreensão da experiência alheia. Começamos descrevendo a análise fenomenológica da empatia de Edith Stein e a relevância contínua de sua crítica às teorias da empatia que pressupõem a necessidade de inferências do "interno" a partir do "externo". Mostramos como teorias prevalentes da empatia, baseadas em tais pressupostos, têm conduzido a confusões conceituais e, em última análise, ao afastamento do fenômeno de interesse. Argumentamos que a noção de empatia implícita nas hipóteses de Skinner sobre como aprendemos a descrever eventos privados se caracteriza por um persistente dualismo interno-externo, que desemboca em incoerência quando levado às suas últimas consequências. Propomos uma conceituação alternativa da empatia, o Modelo Funcional Direto, com três princípios: (1) caráter primariamente dado da experiência, (2) prioridade do todo e (3) interação. Argumentamos que o modelo evita as armadilhas do dualismo interno-externo e oferece um modelo coerente com a filosofia do Behaviorismo Radical.

Palavras-chave: empatia, Edith Stein, behaviorismo radical, teoria da mente, tomada de perspectiva.

Empathy is currently a focus of interest for psychology (e.g., efforts to foster empathy in children, cf. Kucirkova, 2019) as well as for philosophy (e.g., Zahavi, 2014). Although the topic has not been as central in Behavior Analysis, it has been approached through a general model of stimulus control by public accompaniments, sometimes combined with translations of cognitivist concepts of empathy into behavior-analytic terms (e.g., Schlinger, 2009; Spradlin & Brady, 2008). Here, we argue that another account of empathy is possible, one that is more faithful to the functional features of empathic behavior and to the principles of Radical Behaviorism. The framework we will propose (the Direct Functional Account of Empathy - DFAE) is inspired by the philosophical work of Edith Stein, in the phenomenological tradition.

We have organised the paper as follows: first, we introduce Edith Stein's[1] description of empathy, as presented in her work On the Problem of Empathy (1916/1989)[2], especially her discussion and critique of inference-based theories of empathy. We then discuss some difficulties met by current accounts of empathy in the behavior-analytic literature. Finally, we propose an alternative framework: the Direct Functional Model of Empathy. We describe the three main features of the model, which we call (1) Givenness, (2) Priority of the Whole and (3) Interaction. We answer possible objections to each one and describe implications for theory and research.

Empathy

Empathy has many meanings, but in this paper, we will adopt Stein's definition. Stein recognised the plurality of traditions and meanings attached to the term *empathy (*for a recent review of this ongoing plurality, cf. Cuff, et al., 2016), but decided to keep the term because it captures the phenomenon

she sought to describe. By empathy, she meant "the acts in which another's experience is comprehended" (p.6), that is, one´s experience (*Erfahurung*) of another´s lived experience (*Erleben*). A typical example is understanding that a friend is feeling sad. The above definition does not necessarily include acts of sympathy or compassion, but only sufficient demonstration of understanding what the other is going through.[3]

A Very Brief Note on Behavior Analysis and Phenomenology

It is beyond the aim of this paper to defend that phenomenological approaches are compatible with and can bring valuable insights to the philosophy of Radical Behaviorism. For excellent defences of fruitful dialogues between phenomenology and behaviorism, we refer to Willard Day's (1969) now classical paper, as well as Pérez-Álvarez and Sass (2008). We hope this paper will also be an example of the usefulness of exchanges between behavior analysis and phenomenology, through the concrete example of empathy.

Stein´s contributions work at the descriptive level, a level of scientific activity that is prior to experimental analysis. An example might help to illustrate the role of the descriptive level in scientific activity. Let us suppose that a researcher aims to investigate some factors associated with "worrying behavior" in so-called high- and low-worriers. The researcher wants to look at brain activity when people are worrying about important or less important topics. Upon data collection, a participant (previously screened as a high worrier, i.e., someone who worries about various things quite often) arrives, is asked to sit in a quiet room and is interviewed about what he or she worries about. The participant then rates these objects of worry, from less worrisome to most. The worry with the highest score is singled out by the experimenter and the participant is instructed to worry about this topic

---

1 Edith Stein, or Saint Teresa Benedicta of the Cross (1891-1942).

2 All references to Stein in this paper refer to this work, published as a book based on Stein's doctoral dissertation, written under the supervision of Edmund Husserl l (1859–1938) and defended in 1916 (1989 Edition, translated by Waltraut Stein).

3 However, as we will argue towards the end of this paper, the complete separation of empathy from sympathy and compassion is not possible or even desirable once we move away from standard accounts of empathy.

for two minutes, while brain activity is monitored. The same is done for the lowest-rated worry, and in turn this data is compared with the data from other participants classified as low worriers, people who seldom worry about things. The point here is that, even if great care were taken with all methodological steps, the wrong turn had already been taken at the descriptive level, prior to any experimental manipulations. Worrying, in this experiment, was conceived wrongly at the descriptive level, prior to any experimental manipulations, as something that has the property of being started and stopped at will (as operationalized in the instruction to concentrate on worrying about a topic during a time *t*). Control of when to start and stop worrying is not part of the structure of worrying behavior; on the contrary, it often has an involuntary, even intrusive nature. If worrying were something you could start and stop at will, it would not be worrying at all. The subjects, of course, follow the instructions, but what are they really doing? At best, they are simulating their worrying, are remembering them or reenacting them[4]. But a simulated worry is not a worry. In this illustrative example, therefore, the study might produce interesting empirical data, but not about what it was meant to be. It has failed at the descriptive level, even if variable control and statistical analyses have been carried out with great care.

Stein's Aim

Stein aimed to achieve clarity about what empathy is, at the descriptive level, and it is at this level that we think her insights are useful for behavior analysis. She illustrated the priority of phenomenological investigation in relation to empirical research with the following example:

> A friend tells me he has lost his brother, and I become aware of his pain. What kind of awareness is this? I am not concerned here with going into the basis on which I infer the pain. Perhaps his face is pale and disturbed, his voice tone-

less and strained. Perhaps he also expresses his pain in words. Naturally, these things can all be investigated[5]. But they are not my concern here. I would like to know, not how I arrive at this awareness, but what it itself is (p. 6).

Stein describes the various ways in which the other's experience can be given to us. She also discusses how these various ways of comprehending the other's experience express different relationships. However, examining all these different forms of empathy is in itself part of a more general aim: to examine the general structure of how we experience the other's experience. She begins by contrasting empathy with other psychological phenomena, in order to help bring out its distinctive qualities. This strategy also helps to undo various misconceptions about empathy, many of which are still common today.

Empathy and Perception

Most views of empathy in psychology today assume that perception cannot lead to empathy unless it is complemented by inferences about what is going on "inside" the person. According to this view, in empathy, perception does not give us much more than colorless bodily movements, as all the relevant processes happen inside. Thus, the other's emotion, mood or state of mind can only be arrived at through indirect means (by inference or theorizing about what might be going on behind the perceived public manifestation), in contrast to our surrounding physical world, which is perceived directly. The standard account thus supposes that, in perception, the outer world is given to us primordially, while the other's experience is not given to us primordially (needing, for example, to be inferred from public behaviors). We illustrate this common conception in Figure 1.

---

4 Of course, it is possible that the instructions to worry might actually provoke real worrying. But this is contingent and cannot be counted on.

5 The way Stein contrasts the two aims here is slightly misleading, because the two aims are not on the same level (as if they were alternative, equally plausible scopes for her work). The traditional way of framing the problem ('how do we infer state x from outward manifestations?') is incompatible with her view, as will become clear later.
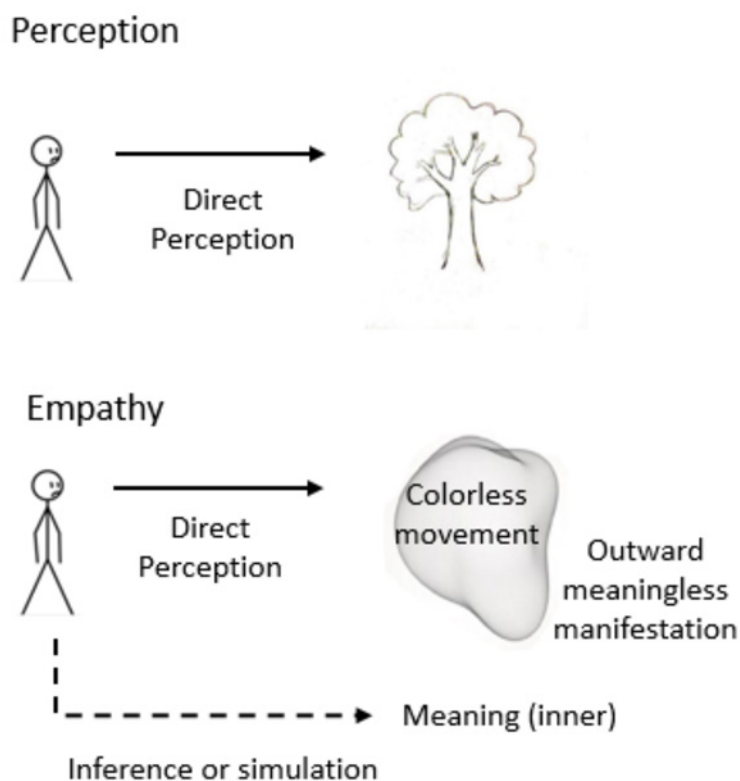
Perception

Empathy

Figure 1: Traditional contrast between perception and empathy.

Stein disagrees with this model of empathy and explains that many things besides the outer world are given to us primordially. The other's experience is one of them. In order to understand Stein's claim and what she means by the primordiality/non-primordiality distinction, it is useful to turn to her second comparison, this time between empathy, on the one hand, and memory, fantasy, and expectation, on the other.

Memory, Fantasy, Expectation and Empathy

An experience is primordial when it is I who experience it – in that sense, all our perceptions, memories, etc. are primordial. But the content of the experience can also be primordial or non-primordial. For example, if someone tells me what a rose smells like, this is a non-primordial experience of smelling a rose. If I smell the rose myself, this is a primordial experience in both senses, that is, it is I who is having the experience, and the content is primordial – it is being experienced directly. However, some experiences are primordial, but have non-primordial content, for example, when I merely imagine smelling a rose, rather than really smelling it. According to Stein, memory, expectation (in the sense of imagining the future), and fantasy all have this in common: they are primordial experiences, but their content is not primordial. For example, when someone remembers seeing a beautiful sunset, the experience of recalling it is primordial, but the content (the sunset) is not. Even as we remember ourselves watching the sunset, this, as it were, "encounter between "I" and "I'"" must keep separate what is primordial (my here-and-now memory experience) from what is not (my watching the sunset). This separateness, plus the non-primordiality of one versus the other, are what make it a memory[6]. This combination of primordial experience and non-primordial content is also present in fantasy (daydreaming), expectations for the future and, as we will now explain, in empathy.

---

6 An example that brings out this point is that of some manifestations of dementia, when the separateness between the primordial experience of memory and the non-primordial experience of the event dissolves.

Three Levels of Fulfilment of Empathy

Stein posits three levels (*Vollzugstufen*) of the fulfilment of empathy. We do not necessarily go through all three levels in every instance. At the most basic level, we grasp the other's experience primordially. For example, I confront my good friend's profound grief over the loss of a loved one. The friend´s grief is right there, in front of me, in all its intensity and heartbrokenness, expressed and available. I experience it directly. The second level can be conveyed through the metaphor of putting oneself beside the other (Svenaeus, 2018), rather than the usual metaphor of "putting oneself in another's shoes." At this second level of empathy, I deepen the first primordial contact with the other´s experience with aspects that I attain non-primordially (for example, I learn that my friend was very close to this person and how important their relationship was. I put myself along her and imagine how this must feel).

Finally, the third level consists of the integration of the previous two. Having perceived my friend's grief primordially and having put myself, so to speak, alongside her, I now face the other´s experience as a whole. I then fully comprehend the other's experience and, this is important, I experience it as, precisely, not mine. This third level of empathy permits me to reflect upon my friend's emotion as it relates to my friend´s context and experience (not mine). For example, upon learning that my friend has been designated to pack away the loved one´s belongings, I imagine how difficult this task must be for her.

It is not always that all three levels of fulfillment occur. For example, if this person is faced with the friend's bare expression of grief but is unable, for one reason or another, to reflect on what the friend is feeling from the friend´s framework of references, then empathy might remain at a level of emotional contagion or the person might not understand the experience because of never having gone through it. When limited to the first level, contact with the other´s experience can lead to feeling overwhelmed, shutting down or seeking escape routes.

We can summarize Stein's position so far as follows: (1) Empathy is a primordial experience that occurs in interaction (e.g., I primordially grasp my friend's grief), (2) with non-primordial content (in our example, the grief never becomes mine, I must be conscious of it being my friend's, lest I fall from empathy into confusion or self-centeredness). (3) The experience points to the other's primordial experience (in this case, my friend's primordial experience of grief).

Why Behavior is not Merely a "Sign" or "Symptom"

Bodily manifestations, such as facial expressions and posture, are often thought to be mere "signs" or "symptoms" of the other's psychological state, with the latter having to be inferred or guessed from those signs[7] (as in Figure 1). Stein disagrees with this and adopts Theodore Lipps's[8] view on the matter. Lipps shows that the notion of a sign does not apply to emotional expressions. He offers an analogy with the well-known situation in which smoke signals fire, in order to show that the relationship between smoke and fire is not the same one that holds between emotional expressions and one's emotions. While smoke is indeed a sign of fire, expressions of sadness, for example, are not merely a sign of sadness. Signs have the feature of pointing to something else, of pointing elsewhere (e.g., the smoke that points away from itself, to the fire). In contrast, expressions are expressions of themselves, so to speak. Emotional expressions already contain the emotion, they are not merely symptoms of something completely different. While smoke, as a sign of fire, takes our attention away from itself (and towards the possible fire), a sad facial expression is not a sign that points elsewhere to some emotion. It is a direct expression of that emotion. Another way of saying this is that a sad facial expression is more like a symbol of sadness (it expresses the emotion) than a sign (that would merely point towards sadness).

7 Cf. the expression `mind reading abilities´, sometimes used in the ToM literature to refer to empathy.

8 Theodor Lipps (1851-1914) proposed the concept of Einfühlung, or empathy, which he described as the act of projecting oneself into the object of perception. The concept was central to his theory of aesthetic appreciation. According to Lipps, the same mechanism of projecting would explain empathy towards another person.

Why Empathy is not a Merging of Selves

Empathy is often seen as a merging with the other. Although Stein agrees with Lipps's arguments about the relation between expressions and psychological experiences, here she strongly opposes his position. Lipps thought that, when we enter a relation of empathy, we temporarily lose ourselves, as we imagine or simulate the other's experience. For Stein, on the contrary, the Self cannot dissolve, even temporarily, if there is to be empathy. Otherwise, one would merely be going through the experience as oneself (as tends to happen in pure emotional contagion). For Stein this is a surrogate of empathy. Real empathy requires the recognition that it is the other's experience that is in question, not mine.

A related point made by Stein is that empathy is not hampered by being in a different state of mind from the other person. As an example, we can think of a parent who is fully absorbed in the joy of a newborn baby. There is no reason, internal to the phenomenon of empathy, why this should hamper empathy towards, say, a friend who is in an opposite emotional state, for example, mourning the loss of a beloved family member. Stein argues that the two experiences are not incompatible, but merely a matter of conflicting objects of attention. The friend's grief might recede to the background when the parent directs attention to the baby, but it would be a strange friend who would say "I am incapable of feeling empathy for you, I am just too caught up in my joy".

What about the common experience of forgetting oneself, the feeling of immersion into the other's point of view? Is this not an important aspect of empathy? Stein mentions Lipps´s example of a circus spectator so enthralled by an acrobat's movements that she drops the program leaflet she was holding in her hand. In this case, the defender of simulation theories of empathy rightly asks, is the spectator not "at one" with the acrobat, as she inwardly follows the acrobat's every movement? Is this not empathy through simulation?

Stein answers that this is a confusion between forgetting oneself and being at one with the other. Again, it is a matter of focus and background, rather than of empathy. If we think about it, we realize that we do not make the same movements as the acrobat[9], even in abbreviated form. It is precisely because we see the acrobat as a separate person that we can follow his or her movements with such fascination.

What about the feeling of oneness that swells up when, for example, people celebrate an event together? Stein´s analysis helps us see that emotional contagion is not equal to empathy and can even go against it. In the example of the group celebration, if sameness of feeling is given too much importance, members of the group might resent someone who feels or expresses feelings differently. Empathy will not mean that everyone necessarily feels the same. On the contrary, empathy will be manifested by an openness to differences in feeling and its expression.

In sum, Stein disagreed with Lipps's idea that the mechanism behind empathy is imitation or simulation, a conception that has been revived in current simulation theories of empathy (e.g., Goldman, 2006). Stein also showed that sameness of feeling cannot, logically, be the basis for empathy. The logical incoherence of taking comparison with one´s one experience as the primary basis of empathy will become evident in the next section.

Why empathy cannot be primarily based on comparisons with one´s own experience

The main rivals to Lipps's conception of empathy as simulation were so-called association theories of empathy, which find their modern corollaries in mainstream ToM (Theory Oj Mind) theories (for an example of a classical paper that originated current mainstream ToM theories, see Premack and Woodruff 1978). Such accounts of empathy assume that the empathizer must guess or theorize what is going on in someone else´s mind, using as "data" the person´s "outer" behavior, allied to their own prior

9 Mirror neurons are sometimes mentioned in this context, but, unless the spectator is herself also a professional acrobat, it is not plausible that there is inner mirroring going on, or at best a very coarse kind of mirroring –the spectator will not have the neuromotor repertoire to "double" the acrobat´s movements on a microscopic scale.

experiences. Thus, according to such accounts, for empathy to happen, there needs to be an "association" between the other´s behavior and representations of similar experiences retrieved from the empathizer's mind. Stein illustrated how association theories work with the example of a person seeing someone else stamping her feet. According to association theory, empathy would happen more or less so: the person seeing the other stamping her feet would associate this perception with a memory of having done this before. The memory being associated with having been angry, an inference would follow: this other person must therefore also be angry.

The mechanism of linking the other´s behavior (stamping) to one's memories or representations, however, cannot explain the empathizer's conclusion ("she must be angry"), because potentially, many other memories could be associated with the observed behavior. For the empathizer to have associated stamping with a memory of anger, and not of dancing to heavy metal or killing a cockroach, at least the first phase of empathy needs to have happened already (the anger needs to have been recognized). In other words, that which the association process is supposed to explain is already presupposed. Logically, the linking of the other´s behavior with one´s own mental representations cannot be the basis of empathy, because linking what one sees with the right memories already presupposes empathy.

Implications for a Behavior-Analytic Model of Empathy

How does Behavior Analysis account for empathy? A simple answer is that empathy is behavior. However, and this is the important point, a good or bad account of what we mean by "behavior" will make all the difference towards achieving a functional account of empathy, rather than a topographical one. Moreover, the most severe problem in accepting a topographical account of empathy is that it mischaracterizes the phenomenon by leading it precisely into the inner-outer dualism for which Radical Behaviorism and its contemporary descendants are supposed to be an alternative.

The model we will propose is "functional" in the broad sense used by Skinner in his Verbal Behavior (1957) and expounded by Willard Day (1969). We mean by it a ground interpretation of empathy that is coherent with how we conceive of organisms, behavior and language. It is our starting point, and if our starting point is dualist, we will only ever get results, in our empirical investigations, of a mentalist conception wrapped up in behavioral jargon. But we are not stuck with either this or abandoning the phenomenon altogether as hopelessly mysterious. A good starting point is possible, one that looks at the sources of control of verbal behavior about empathy, and one that throws light on what we should look for if we want to contribute to understanding this aspect of human interactions. We shall now describe what this starting point might look like. We will do this in two parts: first, we will describe the negative implications of Stein's insights (what they show to be wrong in the standard behavior-analytic view of how we discriminate others' experiences, especially emotions); then we will propose a positive upshot, the Direct Functional Account of Empathy.

Negative implications of Stein's Analysis

The Skinnerian Account of Empathy
Skinner's influential description of how we discriminate private events in others (e.g., Skinner 1945, 1953, 1957) is an implicit model of empathy (although not necessarily on all three levels put forth by Stein). Although Skinner introduces it as an account of how the child learns to discriminate her own "private events", it is also an account of how the other actor involved, the "member of the verbal community", understands what the child is going through and therefore models the right verbal behavior. When we focus on this second actor, the adult trying to understand what is going on with the child (in order to model the right verbal behavior), we see that we are looking at an implicit theory of empathy. In sum, Skinner´s account of the acquisition of verbal repertoires for private events includes an account of how the adult's behavior comes into

a functional relation with the child's behavior, i.e., it includes an account of empathy, as we can see in the following citation:

> The community may resort to public accompaniments of the private event. For example, it may establish a verbal response to an aching tooth by presenting or withholding reinforcement according to a special condition of the tooth which almost certainly accompanies the private event or according to violent collateral responses such as holding the jaw or crying out. Thus, we teach a child to say "That itches" or "That tickles" because we observe either public events which accompany such private stimulation ("the kinds of things which itch or tickle") or some such identifying response as scratching or squirming (Skinner, 1953 p. 259).

According to the Skinnerian account, therefore, X empathizes with Y when X's verbal behavior comes under the control of Y's bodily manifestations and expressions (public accompaniments, e.g., scratching). This will happen *if* X's own public behavioral and bodily manifestations (*a*) have been paired in the past with certain private events (*p*). Since public stimuli *a* and private stimuli *p* together controlled certain verbal behaviors in X, the presence of *a*' in Y is now sufficient to evoke in X the verbal behavior learned under the control of *a* and *p*.

However, if we take this account forward, we will inevitably come against a fatal contradiction. To explain why, let us imagine how we might set up an experiment based on the above account of empathy: (1) create an experimental history of a verbal response V being reinforced in the presence of *a* and *p* occurring together; (2) Test if, upon being presented with just *a*' in another person, the participant would generalize the verbal response V to *a*'.

The obvious problem in this experimental setup is that response generalization can occur without *p* or *p*' ever coming into the picture at all. Simply from the exposition to *a* and having learned the verbal response V in its presence, we could, upon seeing *a*', through generalization, present the verbal response V. The private stimulus *p* seems completely superfluous and, if it is not needed to control either of the answers, then we can leave it out. We have ended up with a description that does away with the link to private events, which is the crux of Skinner's account[10].

The difficulty just described is a consequence of a more general feature of the Skinnerian account: like the Association and Simulation theories scrutinized by Stein[11], *it implicitly assumes that we do not discriminate the other's experience directly*. Although it does not speak of mental representations, and, in this sense, it is not mentalist, on a much deeper level it is a dualist account. It maintains the "inner-outer" distinction and presupposes that, in order for the adult to understand what the child is going through, there must have been some kind of "pairing" between "public accompaniments" and "private events". The term "accompaniment" is very telling: the Skinnerian account of empathy, here, slips into the prevailing conception of behavior as mere accessory, mere sign, mere symptom, that "accompanies" what truly matters. How far from the radical behaviorist principle that behavior is in itself meaningful!

In contrast, the functional model that we will propose assumes that, in standard cases, other people's behaviors are richly informative and meaningful for their own sake. For example, if I know that a friend has just passed a very difficult and important exam with flying colors, and this friend is beaming and jumping up and down, shrieking with delight about the result, what is there still to be unveiled? There is no need to infer or discover anything through simulation, association, 'mind-reading' or what be it. The joy is right there for anyone to see, in full exuberance.

A possible objection to the above example is that even if the joy is very evident, we still cannot feel it ourselves, as that person is feeling it. This is true, but trivial – if I were experiencing it as that person is, then it would become my joy. There would no longer be a relationship involved

---

10 This is very similar to what happened in fact in Okouchi's (2006) experimental simulation of this framework, where, despite several complementary manipulations, some of the participants continuously relied only on the "public" stimulus throughout the experiment (see also Sonoda & Okouchi, 2012).

11 For a description of how similar problems were also inherited by Theory of Mind and Simulation theories, see Leudar and Costall (2009).

and talk of empathy would dissolve. It is precisely because there is a relationship between me and another person´s experience that this is a case of empathy.

Another common objection is that it is possible to simulate joy or to conceal it. This is true, but it is not the standard case, nor the kind we use to teach new members of the verbal community. It would be very difficult for the verbal community to teach newcomers to name anything if they were to start out with such non-paradigmatic, derived cases. For example, it would be hard to teach the concept of sadness by starting out with an example of someone feigning joy. As Skinner (1957) remarks, it is only secondarily that we ourselves learn to conceal or dissimulate our responses. It is also only secondarily that we learn to infer those responses in others in cases where they are being suppressed or dissimulated.

We now turn to the description of the model we propose, the Direct Functional Account of Empathy (DFAE), in which direct discrimination of the other´s experience is considered as the standard case, rather than an impossible feat.

Direct Functional Account of Empathy (DFAE)

We call our proposed functional account "direct" in order to differentiate it from explanations that rely on the idea of indirect access through inference from public accompaniment. The DFAE possesses three complementary features: givenness, priority of the whole and interaction. The three are closely connected, so the reader should bear in mind that each can only be fully understood in relation to the others.

Givenness

The first feature of DFAE, givenness, emphasizes that the emotional experience of the other, in the typical case, is not accessed indirectly through signs or symptoms, but directly. There is no detour into something hidden that must be inferred from public signs, because emotional expression is not merely an outward sign or symptom but the manifestation itself of the emotion, as we exempli-

fied earlier with the case of joy[12]. To give another example, if two people, A and B are discussing a problem and B suddenly stops engaging in the conversation, looks away with an annoyed expression and answers, at best, monosylabically, the annoyance and sudden alienation from the conversation are there to be seen, not as a sign, but as a given. The discriminative aspect of the situation that might spark questions from A is precisely the sudden pulling away and the break in the communication. In other words, if A then asks "What is going on, B? Is something the matter?", it is not because A has no idea what this was an expression of, but precisely because it was so clearly an expression of B's shutting away from the conversation. Even if B then turns back with a big smile and says it was just a joke, the first expression was still correctly identified as distance and annoyance. Just as a spectator watching a play would correctly identify this behavior as an expression of annoyance, all the while knowing that it is pretended annoyance, so B's expression was correctly and immediately identified as annoyance by A, albeit, as A now discovers, feigned annoyance.

One might object that A's discovery that B's annoyance was feigned shows precisely that the partner's private behavior did not correspond to their public behavior and that A was therefore misled by the "outward" behavior. But this is like saying that we are "misled" when we identify an actor's annoyance in a movie, because, after all, the actor is not really annoyed. In fact, the actor's success *depends* on the givenness of emotional expression, on the meaningfulness of behavior. It is because it is given, and not something accessible only through detour, that acting is possible. If annoyance were something private (only signaled by external accompaniments) and we were therefore obliged to make a detour in order to access the private side of the actor's "out-

---

12 Our claim that emotions do not need to be inferred from external aspects, seen as mere signals, is neutral in relation to the current debate in the field of facial expressions, which seeks to answer whether these are intrinsic markers of emotion (e.g. Ekman, Friesen & Ellsworth, 2013) or whether a complex social context is necessary for understanding emotions (e.g. Gross & Barrett, 2011). For this reason, we have not delved into evolutionary theories, nor into recent criticisms of them.

ward" behavior, we would not access annoyance at all, and therefore would never understand the scene.

Another possible objection is that the comparison is unwarranted because the spectators in the theatre know they are watching a fictional scene, while in the case of A and B, one of them was deceived. But this is irrelevant to our point. The aim of our comparison with the theater is to show that the givenness of the other's experience is, precisely, a prerequisite for being able to feign it (as did B) or act it (as did the actor). As Ryle (1949/2009) pointed out, pretending, feigning, acting and the like are "parasitic" on non-feigned, authentic expressions. Their possibility *depends* on the givenness of our expression and on the meaningfulness of behavior.

To show that the previous objection misses the point, we can imagine someone who is deceived by an actor´s behavior. Say the play is Hansel and Gretel (João e Maria) and the spectator is four-year old Ana. Upon seeing a witch angrily and cruelly threaten Hansel and Gretel, Ana is genuinely frightened and starts to cry. Later, it becomes clear that Ana believes the witch is real and has grown wary of her being under the bed or in the cupboard. Ana´s parents comfort her and patiently explain that this was just a play, just pretending, that witches don´t really exist, and that she is safe. Eventually, Ana comes to understand that there is no real witch. However, the fact that the witch was angry and threatening in the story remains true. It was the fictionality of the situation that was not clear to Ana, not the witch's expression or emotion. In other words, and this is the point, after everything is cleared up, the child does not conclude that, in reality, the witch was happy, but that, in reality, there is no witch.

Priority of the Whole

The second feature of the DFAE is Priority of the Whole, and it is closely linked to Givenness. As we have just discussed, people sometimes imagine that, in situations of empathy, we put together the different bodily "signs" of what another person is experiencing, as if it were a puzzle and we were joining the pieces. This is an important, often implicit assumption of many theories of empathy, as we have seen. The DFAE, on the other hand, works with the idea that, although breaking down the other's expe-

rience into "pieces" (this facial expression plus such and such bodily manifestations, etc.) is, of course, feasible, it cannot, logically, be the condition for the recognition of the other´s experience.

Imagining that we recognize what another person is experiencing by putting together loose signs and reaching a conclusion is the kind of horse-before-the cart confusion that Skinner often pointed out. For example, he argued that rule-following is posterior to interaction with the environment and thus cannot be posited as its primary source (e.g., Skinner, 1957). He was not saying that our behavior cannot become rule-governed, but that rules cannot not be the primary, ever-present source of our behavior (as posited by the contemporary computational theories of the mind he was criticizing). Skinner showed that rules are originally derived from behavior, not the other way around. He clarified another cart-before-the-horse confusion when he reminded us that the environment is not naturally cut up into pieces and that we are the ones who neatly categorize it into "different" environments and then are surprised when "transfer of learning" occurs (Skinner, 1938). The same kind of "cart-before-the-horse" confusion is at work when we posit that we empathize by "putting together" discrete "outward behaviors". The analysis of an emotional expression into discrete "parts" can be done, of course, but it is not the basis for the recognition of the expression. One cannot divide something into parts without starting from this something. In an example of this confusion, a lecturer in Behavior Observation Techniques once told his students that they shouldn't say a person was smiling, because this was "inferring", and that they should instead limit themselves to "neutral behaviors". What he seemed to have forgotten is that our behavior would already have to be under the control of the smile in order to choose what "neutral behaviors" to describe. The students had, of course, already recognized the smile as such before they proceeded to dutifully subdivide it into "corners of the mouth going up", "corners of the eyes crinkling", etc. Their recognition of the whole, meaningful action was prior to their effort to transform it into colorless movements.

In the DFAE, therefore, we acknowledge the priority of the whole. Our behavior comes under

the control, for example, of a smile, a friendly gesture, a sad countenance or a sarcastic comment. These are the kinds of functional units that our behavior usually becomes sensitive to, not meaningless twitches and jerks. Here, a comparison with Skinner´s analysis of reading behavior might help. Just as, when reading literature, our behavior comes under the control of whole thematic units and not just formal units (Skinner, 1957) (e.g., the cruelty of a character), so, too, when interacting with others, the functional units that come to affect our behavior are primarily "thematic". To think that they must be meaningless movements that add up in order to control our behavior is to adopt a prejudice about what kinds of functional units can control our behavior.

An immediate objection is that the functional units that control our behavior are a matter for empirical investigation. Should the question of what makes us recognize something as a smile not be an empirical matter? This is correct. However, what is it that we are testing when we set out to experimentally investigate these crucial dimensions? We are testing, of course, the crucial dimensions that make us discriminate *a smile*. So, again, we see the logical priority of the whole (meaningful behavioral units) over its analysis into parts (topographical components, for example).

In sum, we defend that functional analyses of empathy should be open to the discovery of the minimal functional units that control the empathizer's behavior and to the fact that these minimal functional units are, more often than not[13], and certainly in most social interactions where empathy is at stake, things like sardonic smiles, gestures of exasperation and fits of anger, rather than risings of arms, eyelids half closing or a left shoulder rising.

Interaction

In social cognitive accounts of empathy, there tends to be an exclusive focus on the empathizer (largely through the influence of "Theory of Mind" inter-

pretations of empathy, e.g., Baron-Cohen, Leslie and Frith, 1985). The focus of analysis is on how the other's behavior becomes discriminative for the empathizer's verbal and non-verbal behavior. In the DFAE, however, the emphasis is on the *exchanges* between empathizer and empathizee[14], in both directions. This was also Skinner's (1957) approach when he analyzed how speakers edit their verbal behavior to maximize certain responses from the audience (e.g., the autoclitic verbal operants). What we mean is that, in order to understand empathy, the empathizee should not be seen as a static set of stimuli to be discriminated, but as someone whose behavior dynamically affects the probability of empathic behavior (or otherwise) on the part of the empathizer (see Figure 2).

Real-life situations of empathy often involve whole social episodes where the empathizer's behavior not only comes under the control of the empathizee's actions, but also reinforces them differentially. The empathizee thus has his or her empathy-seeking behaviors shaped by the empathizer's reactions, and vice-versa. This interpersonal and interactive aspect of empathy has often been forgotten, because prevalent theories tend to characterize the empathizer as a spectator. The widespread use of the Sally-Anne test [15] (cf. Baron-Cohen, Leslie & Frith, 1985) illustrates the influence of this conception, as noted by Leudar and Costall (2009). The Sally-Anne test models the experimental situation on the assumption that the child is a spectator who must intellectually figure out what is going on between the dolls. The way the test is designed bypasses the well-known fact that empathy, more often than not, occurs in interaction, not in passive onlooking behavior.

In this context, Korkiakangas et al. (2016) conducted conversational analyses of the interactions between experimenter and child during Sally-Anne experiments, showing that said interactions are cru-

---

13 It is of course possible that an idiosyncratic movement might become a "sign" of someone's mood or emotion, e.g., I learn that a certain quick, almost imperceptible pursing of the lips is a sign of contained impatience in my friend. This, however, is not the sense in which traditional theories of empathy employ the term, as explained throughout the paper.
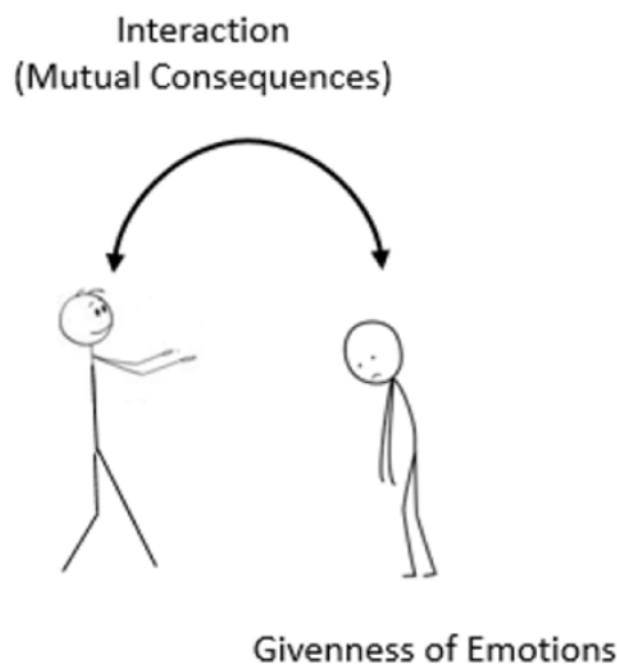
14 We owe this terminology to Berninger (2018).

15 The Sally-Anne test is widely used in developmental psychology as a test of "social cognition" and is closely based on the idea of empathy as a cognitive achievement arrived at by theorizing and guessing the internal and inaccessible mind-states of others, in other words, it is a test based on the inner-outer assumptions that we have been discussing throughout the paper.

Figure2: View of empathy as an interactive phenomenon

cial to understanding the results of the test. Leudar and Costall (2009) had also noted that, although the Sally-Anne Test is supposed to offer evidence for a theory that prioritizes individual cognitive processing and discards social interactions as mere "superficial appearance", the test actually depends on and takes for granted a series of important social interactions between child and experimenter, and that these interactions already include demonstrations of the social understanding that is supposed to be tested.

Recently, Berninger (2018) discussed how real-life empathy, differently from, e.g., empathetic responses towards fictional characters, are often based on complex interactions between empathizee and empathizer. Inspired by insights first articulated by Adam Smith (1822/2010), Berninger pointed out that it is not only the empathizer whose behavior is sensitive to the empathizee's. The empathizee also adjusts his or her manifestations of emotion in a manner attuned to the empathizer's behavior. Berninger (2018) gives the following example of this interaction:

I may be down- right mad when a student shows up for class half an hour late and unprepared. The reasons for my strong emotional reaction might be that I have spent hours preparing for this class and that I care deeply about issues such as mutual respect and punctuality. Nevertheless, I will also be aware of the fact that those present will not be able to share the full-scale of my anger. So, I will do well to exercise some self-restraint and down-regulate my emotion (perhaps only showing signs of irritation). (p. 231).

Berninger's example shows how the whole verbal episode should be looked at when we study empathy, for the final empathic response is a product of the mutual reinforcement between empathizer and empathizee. The professor's (the empathizee's) behavior is not simply a discriminative stimulus for the potential empathizers (the onlooking students) but is also modulated by a history of reinforcement (in this case, differential reinforcement for showing only as much emotion as is likely to be reinforced by the onlooking students).

Behavior analysts have developed useful tools for interpreting functional features of social episodes involving empathy and for tracing how our interpretation of others' experiences is shaped by social consequences during interaction. For example, it is very natural, in the behavioral-analytic framework, to ask ourselves how the empathizee's expressions change and adjust according to the empathizer's behavior, in order to optimize empathic behaviors. We should therefore not limit ourselves to the behavioral "translation" of solipsistic and intellectualist interpretations of empathy that are inconsistent with our basic epistemological tenets[16].

A note on related terms (sympathy and compassion)

Besides avoiding the pitfalls of inner-outer dualism and being much more congenial to the philosophy of Radical Behaviorism, an additional advantage of the DFAE is that it offers an innovative description of the relationship between empathy and sympathy. Many authors have discussed these two concepts, with some defending that they are completely different things (e.g., Chismar, 1988) and others, on the contrary, that they are closely related (e.g., Svenaeus, 2015). The DFAE throws light on this dispute by focusing on how the empathizer and empathizee mutually shape each other's behavior. What we call sympathetic responses (this will also be true for compassion and rational concern) are often manifestations of empathy, i.e., they are evidence that the behavior of one person is sensitive to the other's. For example, when someone shows genuine concern for a friend's suffering and thoughtfully does something to help this friend, these actions show empathy *through* compassion. If the person said, coldly, "I see that you are sad", this would not be a manifestation of sympathy, but it would not be a manifestation of empathy, either. A simple, cold, lifeless labelling of the other's emotion

is not behavior sensitive to the other's suffering. TThe rigid separation between empathy as pure labelling ("cognitive empathy"), on the one hand, and compassion, on the other, does not make sense if we see empathy as behavior attuned to the other. The mentalist notion of empathy as cognitive processing that is separate from behavior ("first I understand [process the input], then I act [output]") is incompatible with a context-sensitive, functional analysis. According to the DFAE, people do not first understand things "inside", then act "outside". Instead, people manifest their understanding of the other through action, which often includes acts of what we call sympathy or compassion.

Final Remarks

Having described Edith Stein's phenomenological account of empathy, as well as her analysis of the problems in traditional theories and their relevance for present-day cognitive theories of empathy, we then showed how some of the same problems are present in the traditional behavior-analytic account of empathy implicit in Skinner´s account of the discrimination of private events.

Overall, we showed that empathy is best seen as a relationship, in which those involved affect each others´s behaviors directly. We suggest that further research might benefit from interpreting empathic interactions with the conceptual tools of the fields of Verbal Behavior and functional interpretation of inguistic acts. This suggestion is in contrast to the current trend of separating "cognitive" "emotional" and "behavioral" so-called "components" of empathy. We must remember that such separations are based on the dualistic roots that presuppose that behavior as such is meaningless in principle. If we are to advance a truly radical behaviorist account of empathy, we cannot adopt as a starting point an acritical acceptance of contemporary theories, such as ´theory theories` (e.g., ToM), that take as a given that we never have access to other people´s experiences or minds, but only to very imperfect signs or symptoms of what is (according to these theories), not contingently, but in principle and forever, private and inaccessible. It is not a coincidence that such theories call the process of empathizing "mind-reading".

---

16   Our model does not claim that this relational structure is ready-made from birth. It develops throughout ontogeny and can be manifested in an incipient way in babies, as has been demonstrated in studies that investigate interactions between phylogenetic and ontogenetic aspects (e.g., Call & Tomasello, 2011; Hoffman, 1984; Hoffman, 2000; Tomasello & Haberl, 2003) as well as cultural issues involved (e.g., Biehl et al., 1997). This broad discussion is beyond the scope of this article.

This article has outlined an alternative path, which, rather than importing frameworks that have profound dualistic roots and superficially trying to transform them into behavioral language, proposes looking at empathy and related phenomena as involving whole behaviors of whole organisms, in this case, organisms who behave verbally, rather than meaningless moving bulks whose experiences must be "guessed"[17] . We elaborated an alternative framework, the DFAE, that (1) does not carry remnants of the dualist supposition that private events are inferred from 'mere' surface accompaniments, (2) accepts that the units that control the empathizer's behavior are, more often than not, thematic wholes and (3) looks at the whole behavioral episode and at how empathizer and empathizee interact and shape each other's behavior.

We are aware that the DFAE is only a starting point, but we believe, with Stein, that starting points are very important for good empirical research. We hope the framework will be improved as empirical findings enter into dialogue with this conceptual proposal and raise the necessity for new distinctions and refinements.

---

[17] Again, we are not saying that we are never in the situation of having to infer or guess other people´s experiences or feelings. However,in theories rooted in the separation between the behavioral and the mental, this is not a contingent possibility, but a principle. Behavior is considered meaningless and only acquires meaning through guesswork about underlying mental events.

References

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21*(1), 37-46. https://doi.org/10.1016/0010-0277(85)90022-8

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal behavior, 21*(1), 3-21. DOI 10.1023/A:1024902500935

Berninger, A. (2018). Empathy Real-Life and Fiction-Based. *Journal of Literary Theory, 12*(2), 224-240. https://doi.org/10.1515/jlt-2018-0013

Call, J., & Tomasello, M. (2011). Does the chimpanzee have a theory of mind? 30 years later. In S. Schleidgen, M. Jungert, R. Bauer & V. Sandow. (Eds.). *Human Nature and Self Design*, 83-96. E-book: Brill.

Chismar, D. (1988). Empathy and sympathy: The important difference. *The Journal of Value Inquiry*, *22*(4), 257-266. https://doi.org/10.1007/BF00136928

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion review, 8*(2), 144-153. https://doi.org/10.1177/1754073914558466

Day, W. (1969). Radical Behaviorism in Reconciliation with Phenomenology. *Journal of the Experimental Analysis of Behavior, 12*(2), 315-328. https://doi.org/10.1901/jeab.1969.12-315

Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings*. Elsevier.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press.

Gross, J. J., & Feldman Barrett, L. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion review, 3*(1), 8-16. DOI 10.1177/1754073910380974

Hoffman, M. L. (1984). Interaction of affect and cognition in empathy. In C. E. Izard, J. Kagan & R. B. Zajonc. (Eds.). *Emotions, cognition,*

and behavior, 103-131. Cambridge: Cambridge University Press.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice.* Cambridge: Cambridge University Press.

Korkiakangas, T., Dindar, K., Laitila, A., & Kärnä, E. (2016). The Sally–Anne test: An interactional analysis of a dyadic assessment. *International journal of language & communication disorders*, *51*(6), 685-702. https://doi.org/10.1111/1460-6984.12240

Kucirkova, N. (2019). How could children's storybooks promote empathy? A conceptual framework based on developmental psychology and literary theory. *Frontiers in psychology*, *10*, 121. https://doi.org/10.3389/fpsyg.2019.00121

Leudar, I., & Costall, A. (Eds.). (2009). *Against theory of mind.* Palgrave Macmillan.

Okouchi, H. (2006). An experimental analysis of another privacy. *The Psychological Record*, *56*(2), 245-257. https://doi.org/10.1007/BF03395548

Pérez-Álvarez, M., & Sass, L. A. (2008). Phenomenology and Behaviorism: A Mutual Readjustment. *Philosophy, Psychiatry, & Psychology, 15*(3), 199-210. https://doi.org/10.1353/ppp.0.0194

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Ryle, G. (2009). *The Concept of Mind (60th Anniversary Edition).* Routledge (Original work published 1949).

Schlinger, H. D. (2009). Theory of mind: An overview and behavioral perspective. *The Psychological Record, 59*(3), 435-448. https://doi.org/10.1007/BF03395673

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* BF Skinner Foundation.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological review*, *52*(5), 270-277. https://doi.org/10.1037/h0062535

Skinner, B. F. (1953). *Science and human behavior.* Macmillan.

Skinner, B. F. (1957). *Verbal behavior.* Prentice-Hall.

Smith, A. (2010). *The theory of moral sentiments.* Penguin. (Original work published in 1922).

Sonoda, A., & Okouchi, H. (2012). A revised procedure for analyzing private events. *The Psychological Record*, *62*(4), 645-662. https://doi.org/10.1007/BF03395826

Spradlin, J. E., & Brady, N. (2008). A behavior analytic interpretation of theory of mind. *International Journal of Psychology & Psychological Therapy, 8*(3), 335–350.

Stein, E. (1989). *On the Problem of Empathy* (Collected Works of Edith Stein, Vol III) [W. Stein, Transl]. ICS Publications.

Svenaeus, F. (2015). The relationship between empathy and sympathy in good health care. *Medicine, Health Care and Philosophy*, *18*(2), 267-277. https://doi.org/10.1007/s11019-014-9601-x

Svenaeus, F. (2018). Edith Stein'sd phenomenology of sensual and emotional empathy. *Phenomenology and the Cognitive Sciences, 17*(4), 741-760. https://doi.org/10.1007/s11097-017-9544-9

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12-and 18-month-olds know what is new for other persons. *Developmental psychology*, *39*(5), 906. DOI 10.1037/0012-1649.39.5.906

Zahavi, D. (2014). *Self and other: Exploring subjectivity, empathy, and shame.* Oxford University Press.