

**Reacting, integrating, and extracting.**  
**A cognitive model of emotion phenomenal mindreading**

The present paper is concerned with emotion mindreading, i.e., the ability to attribute emotions to other individuals. It has three related goals:

- (1) Arguing that emotion mindreading involves both the ability to understand what kind of emotional state another individual is in (*emotion functional mindreading*) and the ability to understand what it is like to be in such state (*emotion phenomenal mindreading*).
- (2) Showing that the available models of mindreading face insurmountable difficulties in accounting for emotion phenomenal mindreading.
- (3) Developing an alternative, empirically grounded model of emotion phenomenal mindreading.

**1.**

Imagine a robot that, when presented with an individual displaying certain behavioral cues, correctly categorizes what kind of mental state that individual is in—where the kind is individuated in terms of its functional role. Nonetheless, it is conceivable that this robot lacks the ability to understand what it is like to undergo a certain mental state.

This scenario illustrates an important distinction between two types of mentalistic attribution: (i) *functional mindreading*, the ability to understand what functional kind of mental state another individual is in; (ii) *phenomenal mindreading*, the ability to understand what phenomenal experience one is undergoing.

In the first section of the paper, by contrasting emotion mindreading with belief attribution, we argue that emotion mindreading involves both the ability to understand what kind of emotional state another individual is in (*emotion functional mindreading*) and the

ability to understand what it is like to be in such state (*emotion phenomenal mindreading*, EPR).

The remainder of the paper deals with the cognitive underpinnings of EPR. In the second section, we argue that the traditional approaches to mindreading fail in accounting for EPR. In the third section, we propose an alternative model.

## 2.

The Theory-Theory (TT) explains mindreading as follows: an internally represented Theory of Mind takes the representation of others' behavioral cues as input and yields a propositional representation of others' mental states as output (Gopnik & Meltzoff 1997; Gopnik & Wellman 1992; Scholl & Leslie 1999; Stich & Nichols 1992).

We discuss two difficulties for TT with respect to EPR:

- The subjectivity problem. Phenomenal experiences are subjective, in the sense that they cannot be understood from a third-person perspective. In fact, to understand them, one has to adopt the subject's point of view (Nagel 1974; Jackson 1982). However, a Theory of Mind is entirely constituted by third-person, objective laws. Therefore, TT leaves phenomenal experience out (Ravenscroft 1998).
- The format problem. We argue that propositional representations do not have the appropriate format to represent phenomenal experience. We consider and reject a possible reply resorting to phenomenal concepts.

The classical alternative to TT, namely the Simulation Theory (ST) (Gallese & Goldman 1998; Goldman 2006; Gordon 1986; Heal 2003), seems to deal with EPR better. Indeed, both the subjectivity and the format problems appear to vanish if we suppose that one attributes emotional experiences by simulating them in one's own mind and then imputing them to other individuals. Moreover, Gallese et al. (2004) argued that a ST-approach to EPR is empirically supported by the discovery of mirror mechanisms (MMs), since they claim that the activation of MMs in an observer generates a conscious experience significantly similar to the one undergone by the observed individual.

We argue, however, that ST faces the following three challenges:

- The subpersonal problem. It is far from clear that the activation of MMs generates any conscious experience at all. However, if the activation of MMs is entirely below the threshold of consciousness, MMs cannot play the role of “bridging first-person experiences” (Gallese et al. 2004, p. 396).
- The lack of sharing problem. EPR does not necessary involve emotional sharing. For example, if one sees a smiling person reaching to pick up a snake, one understands that she is feeling gleeful about it, even though one is experiencing disgust and revulsion (Gallagher 2007).
- The who’s who problem. Even granting that the activation of MMs generates a matching experience, how does the mindreader come to realize that such experience represents the emotional experience of someone else?

As we did before for TT, we consider some replies available to ST, but argue that none proves satisfying.

### 3.

In the final section, we propose an alternative account of EPR. Our starting point is that the aforementioned models oversimplify EPR, since they overlook the complexity of the mindreader’s emotional state. Importantly, we propose that the mindreader’s emotional state is actually the key to EPM.

Here is a sketch of our model:

- (a) The mindreader is in a certain *emotional state* E1, which can be a basic or a complex emotion, an isolated emotional episode or a structured emotional narrative, a phenomenologically vivid or not.
- (b) While being in E1, the mindreader *perceives* another individual displaying some affect-related cues.

- (c) The perceptual state resulting from (b) triggers an *emotional reaction* in the mindreader—call it E2. We specify here the notion of reaction at stake and insist that E2 can be either phenomenologically vivid or not.
- (d) E2 *integrates* with E1, leading to a novel emotional state—call it E3. Drawing on recent evidence, we show why, whereas both E1 and E2 could be below the threshold of consciousness, there are good reasons for thinking that E3 is a phenomenologically vivid emotional state. This addresses the subpersonal problem faced by ST and the subjectivity problem faced by TT.
- (e) The contribution of the observed individual to E3 is separated through interoceptive monitoring processes. Details of this *extraction process* are provided on the basis of general principles and recent empirical evidence (Craig 2002, 2004; Crutchfield 2007; Pollatos 2011). We show why it gives a plausible solution to the who's who problem.
- (f) On the basis of such extraction process, a *phenomenal attribution* is performed. We argue that such attribution is immune to the format problem we raised for TT.