The Fundamental Basis of the Sense of Us

Roberto Horácio de Sá Pereira

Victor Machado Barcellos

Federal University of Rio de Janeiro, Brazil

Abstract

The ability to socialize in early life depends on developing commonalities with others. But what exactly constitutes the "sense of us" or the "we-perspective"? The interaction theory (IT) offers an attractive alternative to mindreading theories, such as theory-theory (TT) and simulation theory (ST), by presenting a further development of enactivism. During embodied interactions, individuals "directly" and "smartly" perceive the mental states of others. Despite the intuitive appeal of direct acquaintance with others' mental states, IT relies on crude metaphors. We aim to retain IT's core intuition—the embodied, embedded, enactive, and affective interaction—within a new framework. First, we introduce Piccinini's key notion of situated neural representations, which fills the gaps that IT leaves behind. Second, we adopt Nicolelis's concept of "brainets". When one animal observes another's behavior, their brains synchronize by simulating the same means to achieve similar goals. We then interpret brainets as situated neural representations: embodied, embedded, enactive, and affective.

Keywords: Mindreading, Theory-theory, Brainets, Simulation theory, Enactivism.

1. Introduction

In humans, the development of social cognition, or the ability to establish social contacts, begins in infancy and relies on building a common basis with others through behaviors such as exchanging gaze, face-to-face interactions, joint attention, etc. As adults, we navigate social interactions based on norms. This challenge, known as social cognition, explores how we comprehend and engage with others in everyday encounters. But what exactly is the primitive "sense of us" or the "we-perspective"?[1]

---

[1] We are grateful to the blind reviewers for their valuable reminder that intersubjectivity encompasses not only the "sense of us" but also the awareness of a shared, publicly accessible world. Our paper, however, concentrates specifically on the fundamental "sense of us".

Argumenta (2025): 1–23 DOI 10.14275/2465-2334/20250.per
ISSN 2465–2334 First published: 18 June 2025

Historically, theories of social cognition have changed dramatically. Early models, such as John Stuart Mill's (1865) analogy theory and behaviorist theories of mind, fell out of favor in the 20th century. The contemporary debate revolves around two main mindreading frameworks: the theory-theory (TT) and simulation theory (ST). TT theorists propose that we understand others by mentally applying psychological theories (Carruthers 1996), while simulation theorists suggest that we simulate the mental states of others through internal simulation models (Baron-Cohen 1995; 2015; Goldman 2006). Nevertheless, this begs the natural question: can infants and animals theorize or even simulate others' mental states, assuming that they have limited cognitive resources? Carruthers, a prominent figure in the TT milieu, believes that they can.

> I am a theory-theorist. I believe that our understanding of mentalistic notions—of belief, desire, perception, intention, and the rest—is largely given by the positions those notions occupy within a folk-psychological theory of the structure and functioning of the mind. […] I also maintain that children's developing competence with these mentalistic notions involves them in moving through a series of progressively more sophisticated theories—for example, moving from desire-perception theory, through a copy-theory of belief, to full-blown, intentionalistic, belief-desire theory (1996: 22).

Finding this kind of answer quite unsatisfactory, Shaun Gallagher and Dan Zahavi, along with others, have advocated for abandoning traditional mindreading approaches in favor of what they term the "second-person approach", also known as the "interaction theory" (IT).[2] This perspective, rooted in phenomenology and the enactivist view of cognition, suggests that we can "directly" and "smartly" perceive the mental states of others without the need for theorizing or simulating. Indeed, recent findings in developmental psychology show that infants can recognize and respond to basic facial expressions and affective states in others. This putative "direct" and "smart" way of perceiving others' mental states forms the basis of social cognition, allowing individuals to engage in meaningful interactions without the need for complex cognitive processes (Gallagher 2001, 2005, 2008a; Gallagher and Hutto 2008; Gallagher and Zahavi 2008, 2011).[3]

As intuitive as the notion of direct acquaintance with others' mental states may seem, it relies on metaphors. Additionally, it assumes a primitive form of sense of us rather than explaining it. But—one may wonder—if our stance towards IT is so critical, why do we devote so much space to it? First, from a historical viewpoint, IT is the first anti-intellectualist alternative to two main mindreading frameworks: the theory-theory (TT) and simulation theory (ST).

---

[2] The literature is vast, and there is no space here to discuss it in detail. See Cassam 2007, Dretske 1973, Gallagher 2004, 2008, 2012, Green 2010, Husserl 1995, Krueger 2012, McNeill 2012, Ratcliffe 2007, and Zahavi 2011.

[3] Indeed, it seems that empathy could be even more ancient in terms of evolution. Evidence concerning tufted capuchin monkeys (Calcutt *et al.* 2017) indicates that they recognize the emotional facial images of their peers faster. Similar to babies, monkeys also demonstrate the capacity to distinguish emotional from neutral expressions. Moreover, when researchers present a specific facial expression (such as anger or happiness), monkeys also actively behave accordingly (*id.*). Finally, there is a significant amount of evidence that chimpanzees, elephants, dogs, and other mammals are also able to empathize (could have some empathy mechanism) (see Custance and Mayer 2012, Plotnik and de Waal 2014, de Waal 2019).

Second, despite its deep problems, we aim to retain IT's core intuition—the embodied, embedded, enactive, and affective interaction—within a new framework. Indeed, Gualtiero Piccinini (2022) did the same with his notion of situated neural representations: he also aims to retain IT's core intuition. Third, we believe that Piccinini's contribution and Miguel Nicolelis's concept of "brainets" fill the gaps that IT leaves behind. When one animal observes another's behavior, their brains synchronize by simulating the same means to achieve similar goals. We then interpret brainets as situated neural representations, that is, embodied, embedded, enactive, and affective.

This paper offers a plausible alternative to interaction theory (IT). While situated within the philosophical domain, it operates at the nexus of philosophy of mind, cognitive science, and neuroscience. This interdisciplinary approach naturally raises meta-philosophical questions that could lead to misunderstandings: namely, how do we conceive the interconnections among these fields? And how are we going to criticize the rival theory (IT)? First, we clarify from the outset that our aim is to establish a new framework for understanding the minimal "sense of us". This undertaking addresses a metaphysical question that is either overlooked in empirical neuroscience and cognitive science or, worse still, yields only vague metaphorical answers when it surfaces.[4]

The paper follows this structure: Section 2 introduces and critiques the standard IT as an alternative to mindreading theories. Section 3 presents Piccinini's framework of situated neural representation. In Section 4, we analyze the brainet framework and support it using Piccinini's moderate situatedness. Section 5 to 7 applies this framework to infants and caregivers, highlighting changes in synchronization areas while maintaining the presence of situated neural representations. The paper concludes with our final remarks.

2. Interaction Theory

The main arguments against traditional approaches on mindreading—TT and ST—may be summarized as follows. The first criticism of mindreading theories of other minds is to overly intellectualize "the sense of us".[5] Additionally, mindreading theories are accused to misinterpreting the cognition of other people's minds as "detached", lacking the essential embodiment and social interaction. Lastly, mindreading theories erroneously presume that theorizing or simulating

---

[4] Nevertheless, despite its traditional nature, we depart from conventional approaches, arguing that this question cannot be answered solely from the armchair or through a priori reasoning. Indeed, we contend that its adequate treatment requires the assistance of neuroscience and cognitive science. Second, we emphasize that we will not employ cognitive science or neuroscience to undermine rival philosophical theories. Rather, we utilize Piccinini's notion of "situated neural representation" and Nicolelis's concept of "brainets" not to rebut IT, but to make sense of what a "direct interaction" might mean. We are grateful to the blind reviewer for begging us to clarity our meta-philosophical viewpoint.

[5] Someone might complain that this criticism (over-intellectualization) applies to TT but is unfair against ST; after all, ST replaces the TT idea of theorizing with the idea of simulation. On closer inspection, though, there is no way of denying that ST's focus is on the cognitive understanding of others. ST portrays understanding others as primarily an intellectual exercise involving mental simulation, imagination, and inference. There is no direct or immediate acquaintance with the mental states of others. This begs the question: how could infants and animals understand each other on such a basis with such poor intellectual resources?

"epistemically mediates" our cognitive access to other people's mental states; intuitively, we do not need any mediation: we "understand" the mental states of others "directly". This criticism sets the stage for the emergence of the enactivist/embodied-based—a.k.a. situatedness—"second-person approach", also known as the "*interaction theory*" (IT).[6] IT proposes that we should acknowledge the embodied and socially embedded nature of psychological life and emphasize immediate, experiential access to other people's minds that precedes and is more fundamental than any imaginative projection (simulation theory) or theoretical inference (the theory-theory).

IT seems appealing for a variety of reasons. First, it breaks with methodological solipsism as a paradigm for social cognition, as found in the traditional mindreading approaches and even in social cognitive neuroscience. IT shifts from studying the cognition of other people's minds and brains in isolation through offline studies to examining social cognition from the first-person perspective of the actors. Moreover, according to IT, the exploration of social cognition takes the form of "*embodied cognition*", where embodied agents interact in real-time with other agents situated in a rich environment.

But perhaps the most appealing and intuitive aspect of IT is the assumption of automatic "*attunement*" between subjects as embodied agents interacting through the exchange of glances, perception of facial expressions and gestures, and joint attention, independent of any reflection, theorizing, or simulation. Following Katja Crone (2020), this can be termed *the sense of us* that underpins what she also calls the "*we-perspective*".

However, this raises the main question: What exactly is this "sense of us"? Against the backdrop of the new phenomenological turn in the philosophy of mind and perception (situatedness approaches), Gallagher answers this question by stating, "We have a *direct understanding* of another person's intentions *because their intentions are explicitly expressed in their embodied actions*" (Gallagher 2004: 224, emphasis added).

Gallagher's crude metaphorical response leaves much to be desired. It prompts the obvious question: What exactly do "understand" and "perceive" signify in this context? Isn't there any difference between "perception" and "cognition?" First, how can one individual "directly perceive"—in a non-metaphorical sense of perceiving—the mental states of another? Second, if the individual in question is a baby or infant who has almost any concepts yet,[7] how can they "understand" (in a non-metaphorical sense of understanding) the mental state of their caregiver? According to Gallagher:

> There are two possible answers. The first would suggest that perception is not good enough to truly *capture the meaning of what the person is doing*. What I perceive directly is that the person's hand is reaching to grasp something or that the person is walking away from someone else. However, perception is not smart enough to recognize that the person is actually getting a drink, or that the person walking

---

[6] In what follows, the terms "enactivism" and "embodied" will be used in a broad sense to accommodate the 4EA framework (embodied, embedded, enactive, extended and affective). We will use "situatedness" to comprise all these concepts.

[7] There is a huge literature regarding the problem of the existence of innate concepts and if babies and infants have any one of them. Although some evidence suggests that concepts such as "number" and "object" are innate, we will not enter into that discussion. See Berent 2020, Carey 2009, Dehaene 2020, Spelke 2022.

> away is doing so for a particular reason, for example, because he is angry […]. A second possible answer is that the perception is smart enough on its own, without the supplement of inference mechanisms, to deliver some sense that the person is getting a drink or that the person is angry and motivated to walk away (Gallagher 2018: 554, emphases added).

As it stands, however, Gallagher only provides negative answers. According to him, the "direct" perception of other states is *not* a "simple" perception of inanimate objects but what he calls a "*smart* perception" in the relevant sense that (i) it "grasps the meaning" of what the person is doing, (ii) dispenses with any additional "inferential mechanism", and (iii) interprets what is seen "in some nonvisual, non-perceptual cognitive steps that go beyond perception itself". Furthermore, (iv) "smart" perception also relies on "pragmatically contextualized comprehension" (Gallagher and Hutto 2008b).

Be that as it may, Gallagher endeavors to elucidate what he metaphorically denotes by "direct" and "smart" perception through an example:

> Therefore, let us distinguish between *a smart perception and a not-so-smart perception* by the following example. *I open my eyes, and I see a certain unrecognized red mass with a specific shape just in front of me*. My eyes are working fine, thank you. My visual cortex is processing all of the preliminary visual information, and what vision delivers is the meaningless red mass, which *I then have to interpret in some nonvisual, nonperceptual cognitive steps that go beyond perception itself. Let us call this a not-so-smart perception*. In contrast, in the very same situation, when I open my eyes, I see my car. It is true that it has a specific shape and is red, and I do see the shape and the color, *but I see the shape and color as being aspects of something that is amazingly recognizable as my car*. Actually, if you ask me what I see, I would likely not say that I see a red and shapely mass. *Somehow I see through those aspects and I see my car. I do not see red mass, shape, and color, and then try to piece all of that together to make it add up to my car. I simply and directly see my car*. Let us call this a perception with some degree of smarts (2018: 554, emphases added).

Pierre Jacob's (2011) observation presents a significant challenge to IT. He argues that one can only recognize something "as my car" if one possesses the *necessary background conditions*. In essence, IT faces a dilemma: either it must embrace a kind of disguised behaviorism or acknowledge that understanding another's goal or emotion from their behavior is *not direct* but requires contextual assumptions. "Smartly" perceiving of shapes and colors as of "my car" necessitates background knowledge, contextual cues, past experiences, and relevant concepts such as color concepts, shape concepts, the concept of a car, and self-concepts. Without these elements, what remains is a nonconceptual content of "not-so-smart perception" (Jacob 2011: 519). This critique remains even if we assume that babies have some core concepts, such as the concepts of "number" and "object". It is reasonable to assume, as developmental psychology seems to show, that babies don't have complex concepts such as "car" and "cloud".

Nonetheless, Gallagher adds: "Phenomenological approaches to cognitive science insist on the primacy of a person's embodied, interactive, and directly perceptual (i.e., not theoretical or explicitly conceptualized) grasp of another's mind" (2012: 437, emphasis added). However, the notion of "smart direct perceptions" remains crudely metaphorical and even ambiguous, particularly regarding whether "smart perceptions" are genuinely direct perceptions or rather "direct perceptual judgments" (see Block 2023: 11–13, 33, 42–44). All these obscurities make it almost impossible to understand the key claims of IT that preverbal infants, and toddlers engage in this kind of smart and direct perception.

Regarding this common old objection (behaviorism), Zahavi (2011: 549) complains that it is *prima facie* odd to accuse the enactivist-based IT of behaviorism since IT traces back to the phenomenological tradition in the first place. The most salient feature of phenomenological tradition has been its recurrent emphasis on the importance of subjectivity and of the first-personal perspective. In contrast, old behaviorism is a form of reductionism of the first-person perspective to behavior, which is a third-person aspect of a subject experience.[8]

Enactivist-based IT does not endorse reductionism in the sense of ontological behaviorism, which attempts to equate mental states with behaviors or dispositions thereof. Instead, IT recognizes only that bodily expressiveness plays a crucial role in understanding mental life but does not seek to reduce mental states to behaviors or dispositions. Zahavi argues that the question is not whether bodily expressiveness exhausts our mental life but whether it possesses intrinsic psychology (*id.*). The question is whether there is a conceptual or essential connection between bodily expressiveness and mental states, or, as Zahavi puts it, if "bodily expressions are already from the start soaked with mindedness" (*id.*).

However, Zahavi's phrase "bodily expressions already from the start soaked with mindedness" is replacement of one crude metaphor with another. The precise meaning of this expression is not entirely clear, and it raises the recurring issue of whether bodily expressions inherently possess "mindedness" without the involvement of background knowledge, contextual clues, previous experiences, and concepts.

Suppose that you and I are embodied interacting when I perceive that your hand is reaching to grasp something (Gallagher 2018: 554). That is nothing but bodily movement. Now, let us assume for the sake of argument that I perceive and understand that you are trying to reach something (the refrigerator door handle) (*id.*). The problem is that such a bodily movement is compatible with many quite different intentions. For instance, it is compatible with the intention "the person is actually getting a drink" (*id.*) but also with the different intention of closing the refrigerator door, with the intention of investigating whether the fridge is in working condition, with the intention of grabbing something to eat, or with the intention that there is a dead human body in the fridge and so on *ad infinitum*. I can only choose one of those excluding intentions if I know the background conditions, the contextual cues, past experiences, etc. For example, knowing that we are drinking and chatting together, I tacitly assume that you are trying to get another beer. But again, in this case we are back to the second horn of Jacob's dilemma: there is no direct access.

3. Situated Neural Representation

The radical enactivist trend in the milieux advocates for the elimination of the core notion of content. The relevant contributions here are made by Daniel Hutto and

---

[8] Or the elimination of the first-person view/consciousness from the practice of a serious and rigorous scientific psychology. See Watson 1913.

Erik Myin (2012, 2017).[9] They defend what they call the "radicalized" version of "enactivism"—REC. The leitmotif is "going wide!" (i.e., to appeal to world-encompassing interaction processes in cognitive science explanation). It is no longer just a matter of dispensing with the concept of mental image but even with what we now understand as representational content. They claim that "The activity of the organism—engaging with features of its environments in specific ways—is sufficient for the most basic types of cognition. Such activity does not depend on individuals retrieving informational content from the world" (2012: 4–5).

However, not all authors who endorse certain versions of the situatedness framework seek to reject or eliminate the crucial role of representations in explaining certain types of behavior.[10] One of these authors is Gualtiero Piccinini (2021, 2022, forthcoming). In our view he has a more nuanced an insightful view than radical enactivism. As he exposes:

> Far from being opposites, situatedness and neural representation are more deeply intertwined than anyone suspected. I introduce *a neurocomputational account of cognition that relies on neural representations*. I argue not only that this account is compatible with (non-question-begging) situated approaches but also that it requires embodiment, embeddedness, enaction, and affect at its very core (2022: 1, emphasis added).

We cannot argue in favor of Piccinini's view in this context for a question of space. Instead, we will present his insightful perspective and demonstrate how it restores the indispensable notion of representational content within the framework of moderate situatedness cognition. We believe that his approach provides the initial step in clarifying the "metaphors" and addressing the gaps of standard IT, establishing the new basis for IT in the following section. In other words, what we propose is not a rejection of IT, but a modified version of it that incorporates a notion of representational content compatible with the 4EA framework. The specific notion of representational content that we appeal is Piccinini's situated neural representation.

But what is a situated neural representation (SNR)? Primarily, they are structural representations, serving as models of a goal that can guide behavior in relation to that goal insofar as they display four main characteristics. Firstly, they are homomorphic to their target; that is, they are 'partially isomorphic' to their target (Piccinini, 2022: 4). Secondly, they are activated by stimuli emitted by the target. Thirdly, they can control behavior to achieve that goal. Finally, they can decouple from the stimuli emanating from the target, allowing them to control behavior in an offline way that achieves the goal even when the target no longer directly activates the representations (Piccinini 2022: 4–5). To be clear, A is partially isomorphic (homomorphic) to B if A maintains structural similarities with B. When we talk about the relation between the structural representation R and a state of affairs P, we are not claiming that R has all the same first-order properties as P, but only their second-order relations. Obviously, the structural representation R of an

---

[9] For more relevant literature on radical versions of enactivism/embodied cognition, see Chemero 2009, Di Paolo *et al.* 2018, Myin 2016.

[10] Andy Clark (1997) is one of the most notable philosophers who tries to reconcile representationalism with embodied cognitive science. Some authors interpret Alva Noë (2005) with a moderate enactivist position similar to Clark's view.

apple on the table does not possess properties such as being red,[11] rounded, etc. In other words, there is nothing red or rounded in our brains when we represent an apple. For instance, consider a neural map in a monkey's brain that represents its room with a dispenser full of grapes and orange juice. As a neural map, the structural representation is homomorphic to its target. The representation builds an internal simulation of the environment that is partially isomorphic to it, is activated by signals coming from the target, is capable of controlling behavior to achieve the goal/target, and is able to decouple from the target if it no longer directly activates it. The monkeys could even navigate using representations without external input, using only previously recorded information. For example, the monkey could know through its representations of grapes that there are grapes in the dispenser even if the grape's sensory stimuli are not present to its visual system (Neupane, Fiete and Jazayeri 2024).

Piccinini highlights the importance of considering the representational content when defining a structural representation. He points out that many theorists do not differentiate between the *functional role* and the *representational content of structural representations*. Piccinini (2021, 2022) suggests that Informational Teleosemantics (Dretske 1995; Neander 2017) provides the best explanation of the content of structural representations in relation to the function they acquire in conveying information about their target.[12]

However, the structural representation has not only the function of carrying information about its target. It has the additional function *simulating* its target, *by constructing and sustaining predictions of both the subject's body and the environment*. It utilizes this simulation to regulate behavior through motor commands and continuously updates the simulation with information from the body, environment, and its own motor commands (Piccinni 2022). It works similarly to a generative model, constructing a representation of a certain state of affairs by predicting the hidden causes of sensory data and updating this model through corrections based on prediction errors from perception and bodily activity (see Clark 2015).

Based on the considerations presented, neural representations require a neurocognitive system that is embodied, embedded, enactive, and affective (Piccinini's central claim). This "situatedness" is absolutely essential because neural representations, along with their associated computations, develop over time through the dynamic interaction between the nervous system, the body, and the environment, aligning with the organism's needs. Therefore, situated neural representation is embodied, embedded, enactive, and affective. Let us delve deeper into Piccinini's proposal.

Neurocognitive systems are different from traditional AI systems because they continuously adjust their bodies to act within the environment. They use efference copies of their motor commands to predict changes in sensory inputs. Additionally, they collect data on the effects of their motor commands on both their body and the environment immediately after implementing these commands. By utilizing continuous, real-time feedback, neurocognitive systems refine their structure and improve their overall performance. In other words, unlike

---

[11] Indeed, the brain has a red color because of the blood circulating in it. But the point is that the brain does not has the color of the *representatum*.

[12] For the purposes of our argument, it will not be necessary to endorse Informational Teleosemantics or any other theory of mental representation. We will remain neutral in the debate over whether this is the best naturalistic theory of mental representation.

traditional computers that generate outputs solely based on their inputs and internal states, neurocognitive systems are also capable of active learning. They combine various sources of information with their self-organizing capacity to modify their structure and, as a result, their future functions.

What does active learning entail? First and foremost, active learning necessitates *embodiment*, meaning a close, dynamic connection between the neurocognitive system and the body. This is important not only for the obvious reason that the body houses the sensors and effectors required for the neurocognitive system to receive and act on information but also because the real-time feedback loop between the neurocognitive system and the body is crucial. In this loop, the body moves in response to neural activity and quickly relays information back to the neurocognitive system about these movements. This feedback is essential for the neurocognitive system to learn how to represent its body, distinguish it from the environment, and effectively simulate and control its movements. In summary, the neurocognitive system cannot reach its full learning potential or learn to control its body within its environment using internal simulations as a guide without continuous, dynamic interaction with its own body.

Active learning also involves *embeddedness*, which is a close and dynamic interaction between the nervous system, body, and environment. This interaction is vital not only because the environment provides the primary sources of information for most senses (except proprioception) or because the body relies on its environment to function properly. Furthermore, the real-time feedback loop between neurocognitive systems and their environment is crucial for learning how to represent the environment, differentiate it from the body, and effectively simulate and act within it.

Active learning also involves *enaction*. Piccinini describes "enaction" as a dynamic interdependence between a cognitive system and its environment that develops continuously in real time. This means that cognitive states and processes influence the organism's body and environment, while the body and environment, in turn, impact cognitive states and processes. Each motor command affects (a) the movement of the body, (b) changes in sensory input (due to alterations in the body's position relative to the environment), and (c) changes in the environment (as the organism's actions alter it). Additionally, the simulation attempts to predict how these factors will evolve, and the system compares these predictions with its sensory data.

Finally, active learning necessitates *affect*. Piccinini describes "affect", in its simplest form, as a system of internal signals that assess the state of the organism and its environment to guide the selection of actions that meet the organism's needs. For complex animals with behavioral flexibility, affect is essential for choosing actions that fulfill their needs, assessing external situations, and eventually learning to select adaptive action sequences for various situations. As discussed, affect in this context is a component of reinforcement learning, which, in turn, is a part of active learning.

4. Brainets

In his book of 2020, Nicolelis uses the term "brained" for the first time to refer to a conceptual framework for a network of interconnected brains. The research explores the possibility of linking multiple brains, first, through brain-computer interfaces and then, through a direct brain-to-brain communication. The goal of this "brainet" is to create a system where information can flow between individuals' brains, potentially enabling collaborative problem-solving and a new understanding of collective cognitive processes. This idea represents an ambitious step towards integrating neuroscience and technology to enhance inter-brain communication and social collaboration.

In their early paper of 2012, Uri Hasson and colleagues propose that this binding mechanism reflects the coupling of action and perception within a single cortical system. Different objects in the environment emit different forms of energy, which receptors convert into electrical impulses. The brain then utilizes these signals to extract information about the state of the world and generate appropriate intentional behaviors (Hasson *et al.* 2012: 2). The figure below illustrates two types of coupling: (a) stimulus-to-brain coupling and (b) brain-to-brain coupling.

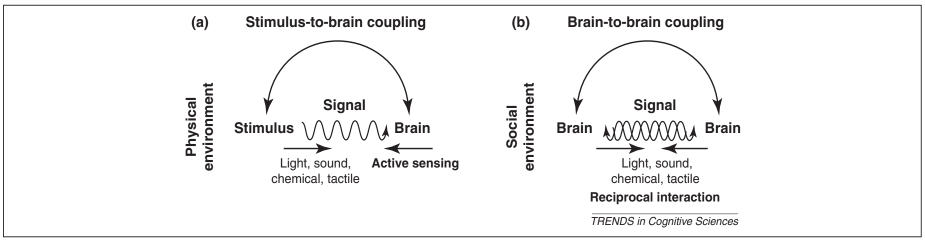


Figure 1. Illustration from Hasson *et al.* 2012: 115

In a noteworthy experiment known as the B-3 brainet, researchers developed a computer system capable of amalgamating the raw electrical activity simultaneously generated by three individual monkey brains. Notably, each monkey was situated in a different location and had no direct physical contact with the others. In this experiment, the researchers tasked each of the three monkeys with utilizing their brain's electrical activity to control the movement of an avatar arm toward a target within a 2/3D virtual environment. Successful completion of this task resulted in a reward, typically a delicious fruit juice. However, each monkey could only control two of the three coordinate dimensions of the virtual environment (X, Y, and Z). For instance, monkey A could control the virtual arm movement along the X and Z dimensions, while monkey B could control the X and Y dimensions. The computer's role was to amalgamate all the inputs derived from the brain and produce the outcome as a three-dimensional arm movement. Crucially, the precondition for success was as follows: The synchronization of the motor cortexes of each monkey was essential to generate a three-dimensional movement of the virtual arm toward the target (Ramakrishnan *et al.* 2015, Nicolelis 2020). In essence, their motor cortexes coupled their brains in a "brainet".

After initial trials in which the monkeys could easily control their two-dimensional arm coordinates, they faced challenges in achieving synchronization in their motor cortices, preventing them from performing three-dimensional arm movements in the virtual environment. However, after three weeks, they successfully performed 3D arm movements. Remarkably, the motor cortex of each monkey had learned to synchronize and cooperate, enabling the movement of the avatar arm. The data from this experiment showed a correlation between successful trials of 3D arm movements and increased levels of cortical synchronization in the monkeys' brains. Groups of cortical neurons in each monkey's brain began to show electrical activity simultaneously, leading to the formation of a brainet (Ramakrishnan *et al*. 2015, Nicolelis 2020).

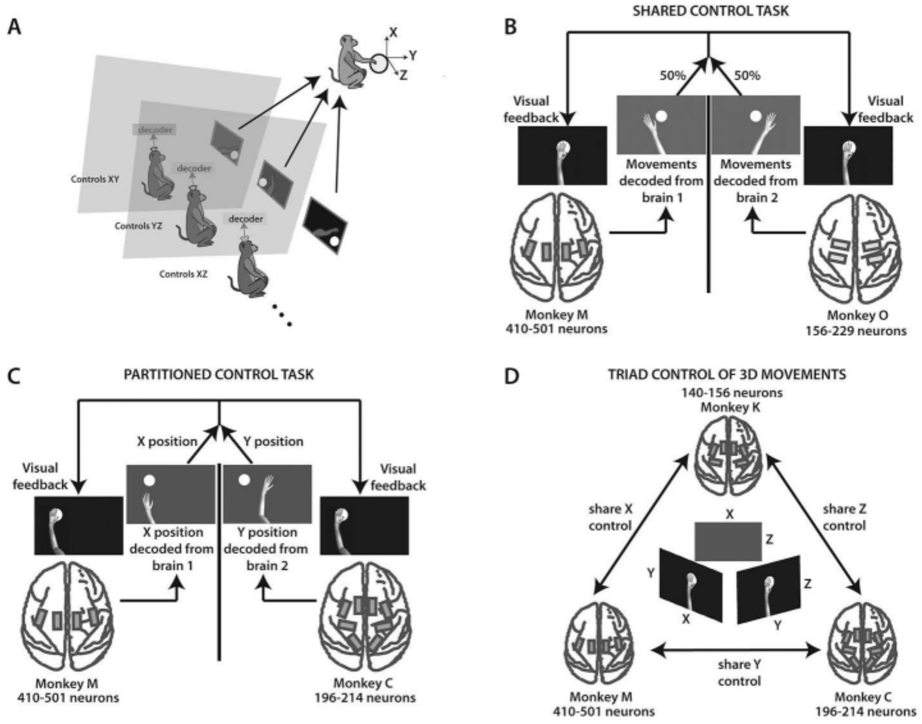


Figure 2. "Different configurations of monkey brainets: A: General arrangement of a monkey brainet used to perform a shared motor task. The monkeys were located in separate rooms. Each monkey faced a computer screen that displayed a virtual avatar arm. The behavioral task consisted of using three-dimensional movements of an avatar arm to reach virtual targets on a screen. The 3-D avatar arm movements were produced by the combination of the cortical electrical activity produced simultaneously by the group of monkeys forming a given brainet. B: An example of a shared motor control task in which each of two participating monkeys contributed 50% to the (X, Y) position of the virtual arm. The cortical locations of the implanted microelectrode arrays are shown below the task diagram. C: Partitioned control task, in which one monkey contributed to the X position of the avatar arm and the other to the Y position. D: Detailed representation of a three-monkey brainet task. Each monkey performed a two-dimensional task, and all three together controlled three-dimensional movements of the avatar arm" (Ramakrishnan *et al*. 2020: 132)

In 2018, Po-He Tseng and colleagues conducted an even more remarkable experiment, known as the "passenger" and "observer" experiment, in Miguel Nicolelis' laboratory. In this experiment, monkeys could see each other instead of interacting through computer interface. The study involved simultaneous wireless recordings of cortical neural ensemble activity from pairs of monkeys to investigate the neural correlates of spatial social interactions in primates. The setup was as follows: Two monkeys (the so-called "passenger" and the so-called "observer") were placed in a 5.0-by-3.9 m room. The "passenger" sat in an electrically actuated wheelchair, while the "observer" sat in a stationary chair placed in the corner of the room. During each trial, the passenger moved from a starting location (shown on the left) to a stationary grape dispenser. This task was a socially cooperative one with a mutual benefit goal. If the "passenger" successfully navigated the car to the grape dispenser and picked up the grapes, the "observer" received orange juice as a reward for its partner's success. Consequently, the "observer" paid close attention to every movement of the "passenger".

Paraphrasing the explanation provided by Tse and colleagues (2018), the picture below depicted five representative wheelchair routes in different colors, with a computer program randomly generating these routes. (C) Color diagrams show the neuron ensemble's activity in two representative trials (Tse *et al.* 2018: 2). Each horizontal line corresponds to one unit, with the color representing the normalized (z-weighted) firing rate of 69 units recorded in monkey C (the observer) and 47 units in monkey K (the passenger) (*id.*). Red horizontal lines indicate instances of synchronization episodes of brain-to-brain coupling (BBC). (D) Continuous BBC analysis for the trials shown in (C) was conducted (*id.*). The researchers calculated the instantaneous distance correlation values using a sliding window with the same 3-second width as the red bars in (C) (*id.*). Arrows indicate the peak values of the correlation (*id.*). (E) The wheelchair routes correspond to the same trials as those in (C) and (D) (*id.*). The routes are color-coded to indicate BBC (*id.*).

After several failed attempts, both monkeys eventually learned how to accomplish what the researchers had in mind. The following occurred during the task: The passenger maneuvered their robotic chair, and the observer tracked them with their eyes. At this juncture, the mid-premotor cortex (PMC) and the dorsal premotor cortex (dPMC) of each monkey's brain impressively increased their electrical activity simultaneously, indicating that these brain regions in each monkey concurrently *represented* the same means (driving the wheelchair to the grapes) to achieve a similar goal.

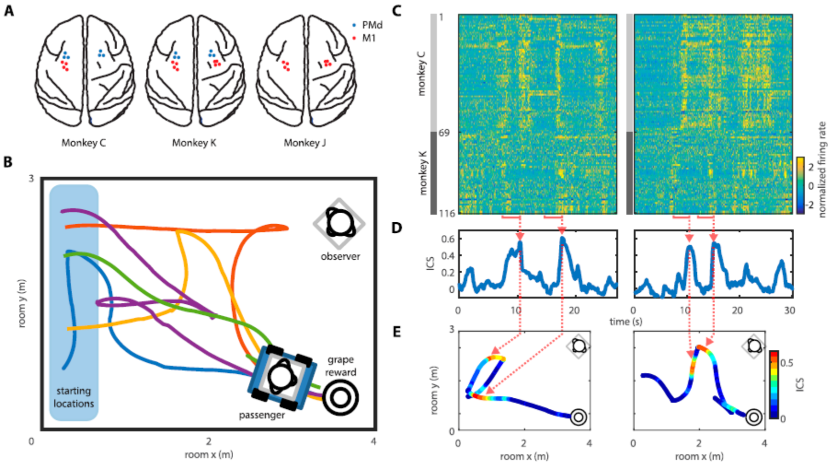


Figure 3. Illustration from Tse *et al.* 2018: 2

In slightly different words, it was discovered that both the PMC and dPMC groups of neurons of the passenger and the observer modulated their firing rate according to their performance and movements (position and speed) of the wheelchair. More importantly, the groups of neurons recorded in their brains "exhibited episodes of transient synchronized firing" (Tseng *et al.* 2018: 7). Therefore, they have their brains coupled together, forming a brainet, a brain-to-brain connection.

Another interesting finding was that the cortical synchronization found in the brains of each of the monkeys (the passenger and the observer) behaved in a special way when the two were close to each other (approximately three feet apart). That is, during a context of social interaction between the two monkeys (the passenger and the observer), the electrical activity produced in the PMC of each monkey also occurred simultaneously. Moreover, the researchers also found that the synchronization of groups of neurons in the PMC of each of the monkeys during social interaction relies on the social hierarchy between them: they show significant changes in electrical activity in the PMC/dPMC when the dominant monkey was the passenger and when he was the observer. Given only the neural activity in the monkeys' PMC/dPMC, the researchers were able to fairly accurately demonstrate which monkey was dominant and dominated (Tse *et al.* 2018).

5. SNR as the Basis of Brain-to-Brain Coupling

This paper's central question is about the metaphysical nature of the shared neural network. The answer lies in Piccinini's concept of SNR. Let's explore this step by step. To begin with, the share neural network must take the form of a structural representation, displaying the four key features. Firstly, they are homomorphic to their target. Secondly, they are activated by stimuli emitted by the target. Thirdly, they are able to control behavior to achieve that goal. Fourthly, they can decouple from the stimuli emanating from the target, allowing them to control behavior in a way that achieves the goal even when the target no longer directly activates the representation.

Let us reconsider the case of the passenger/observer. First, they share a neural pattern that is homomorphic to their target, such as a map of the room where the "passenger" sat in an electrically actuated wheelchair. Second, the neural pattern is activated by the same target, such as the dispenser full of grapes and orange juice. Third, the shared neural pattern can control behavior to achieve their goals. Fourth, the shared brain pattern can be decoupled from the stimuli emanating from the target.

However, as we have discussed, the simulation of the goal depends on the ongoing simulation of both the environment and the bodies involved: the acting body and the stationary body. It utilizes this simulation to regulate behavior through motor commands and continuously updates the simulation with information from the body, environment, and its own motor commands. This process is precisely what Nicolelis (2020) describes in instances of failed trials. For example, each time the passenger attempted to perform the task after the start signal and failed, the passenger's brain "immediately remapped the new location of the grape dispenser as well as its relative spatial position to the wheelchair under the passenger's brain control" (Nicolelis 2020: 134).

But where does this lead us? To the idea that the synchronized cortical areas (the electrical activity produced in the PMC of each monkey also occurred simultaneously) are nothing but *situated neural representations*. First and foremost, brainets require *embodiment*, i.e., dynamic connection between the neurocognitive system and the body. This is important not only because the body houses the sensors and effectors required for the neurocognitive system to receive and act on information but also because the real-time feedback loop between the neurocognitive system and the body is crucial.

The synchronized cortical areas, where the electrical activity generated in each monkey's PMC occurs simultaneously, are situated neural representations for a second reason: *they are embedded within an environment*. This is significant not only because the environment is the primary source of information for most senses, and the body depends on it to function properly, but also because the real-time feedback loop between neurocognitive systems and their environment is crucial for accurately representing the environment, distinguishing it from the body, and effectively simulating and interacting with it.

The synchronized cortical areas are situated neural representations for a third reason: *enaction*. Each motor command affects (a) the movement of the body, (b) changes in sensory input (due to alterations in the body's position relative to the environment), and (c) changes in the environment (as the organism's actions alter it).

Finally, the synchronized cortical areas are SNR because they are affective. As animals with behavioral flexibility, primates rely on affect to choose actions that fulfill their needs, assess external situations, and ultimately learn to select adaptive action sequences across various contexts; for example, the respective rewards: grapes for the passenger and orange juice for the observer. It involves the satisfaction of biological needs that provokes experiences with valence—positive or negative—which typically characterizes affective states (Birch 2024, Carruthers 2024).

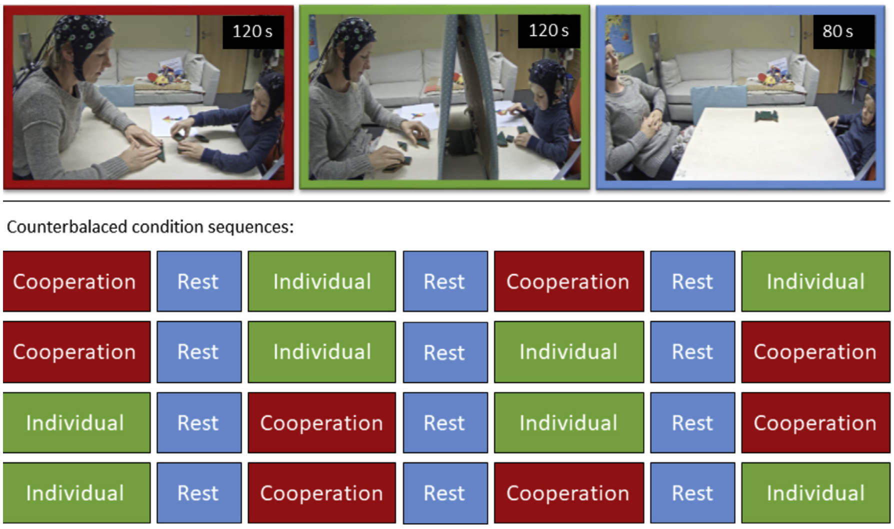
6. Infants and Caregivers

Below, we will present some important empirical research that deals with the brain-to-brain coupling (BBC) between caregiver and the baby. The aim here is not to provide an exhaustive discussion of all the available evidence, but to present key evidence relevant to the hypothesis we are proposing.

In the Po-He Tseng and colleague's (2018) experiment, interbrain synchronization was observed in the PMC (mid-premotor cortex) and the dPMC (dorsal premotor cortex) for the obvious reason that the Po-He Tseng and colleague's experiment was about to drive a robotic chair. This finding is congruent with IT since for enactivism "perception" or "understanding" relies on embodied actions and embodied interactions. However, as what is at stake is now the BBC between infants and caregivers, we cannot expect to find the BBC in PMC and dPMC, but in different cortical areas.

Jiang and colleagues (2012) found a significant increase in synchronization in two brain's *left inferior frontal cortices* during a face-to-face dialog but none during different communication conditions, including during a face-to-face dialog between partners but none during a back-to-back dialog, a face-to-face monolog, or a back-to-back monolog. When two people establish a face-to-face communication, their brains become coupled; that is, their brains show electrical activity that produces the left inferior frontal cortices synchronously. Jiang and colleague's experiment was carried out by means of fNIRS hyperscanning (fNIRS measures change in regional cerebral blood flow by quantifying the changes in oxyhemoglobin concentration) (Jiang *et al.* 2012).

In a recent paper (2020), Trinh Nguyen and colleagues studied interpersonal neural synchrony using the same fNIRS hyperscanning. They examined the effects of interaction quality on neural synchrony during a problem-solving task in 42 dyads of mothers and their preschool children. Their results revealed increased neural synchrony *in the bilateral prefrontal cortex and temporal-parietal areas* during cooperative tasks compared to individual problem solving. Moreover, higher neural synchrony during cooperation seems to be correlated with higher behavioral reciprocity. They then conclude that “neural synchrony is a biomarker for mother-child interaction quality” (Nguyen 2020: 235). The figure below summarizes their empirical findings.



Counterbalaced condition sequences:

Cooperation	Rest	Individual	Rest	Cooperation	Rest	Individual
Cooperation	Rest	Individual	Rest	Individual	Rest	Cooperation
Individual	Rest	Cooperation	Rest	Individual	Rest	Cooperation
Individual	Rest	Cooperation	Rest	Cooperation	Rest	Individual

Figure 4. “Study setup during cooperation (left) and individual (middle) problem solving as well as the rest phase (right). Rows indicate possible sequences to counterbalance (Latin square) the order of tasks” (Nguyen et al 2020: 238)

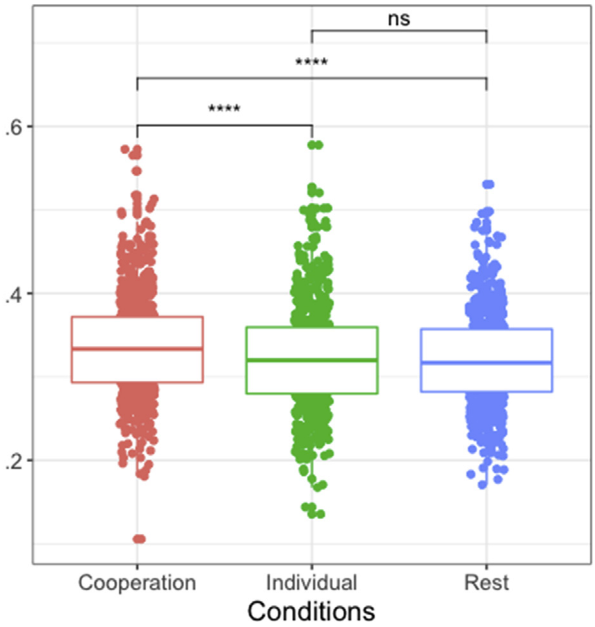


Figure 5. “Plot of the main effect of condition. Neural synchrony during cooperation was significantly higher than during individual and resting phases (averaged across all ROIs). ns 1⁄4 nonsignificant, **** 1⁄4 p <.0001” (Nguyen et al 2020: 241)

Elise Piazza and colleagues (2020) have recently undertaken the most telling study about the BBC between babies and caregivers. Studying real-life, face-to-face communication between babies and adults is quite difficult. Indeed, most studies of neural coupling were conducted in Hasson's lab and in Nicolelis's lab with a functional magnetic resonance imaging (fMRI) method, involving hyperscanning adult's brains laying down and watching movies or listening to stories. However, to study real-time communication with babies and preverbal infants, the researchers needed to create a child-friendly environment and method of recording brain activity simultaneously from baby and adult brains. Therefore, Piazza and colleagues developed a new dual-brain neuroimaging system that uses functional near-infrared spectroscopy (fNIRS), which is highly safe and records oxygenation in the blood as a proxy for neural activity. The new setup allowed the researchers to record the neural synchronization between babies and an adult's brains while they played with toys, sang songs and read a book (Piazza *et al*. 2020).

The same adult interacted with all 42 infants and toddlers (N=18; 9-15 months of age) who participated in the study. Of those, 21 had to be excluded because they "squirmed excessively", and three others flat-out refused to wear the cap, leaving 18 children, ranging in age from 9 months to 15 months. The experiment had two parts. In the first, the adult experimenter spent five minutes interacting directly with a child—playing with toys, singing nursery rhymes or reading Goodnight Moon—while the child sat on their parent's lap. In the second, the experimenter turned to the side and told a story to another adult while the child played quietly with their parents. The caps collected data from 57 channels (3 cm in the adult, 2.5 cm in the infants) across the cortex of each participant. These channels covered the prefrontal cortex (PFC), temporoparietal junction, and parietal cortex (areas involved in prediction, language processing, and understanding other people's perspectives). The locations of these channels were homologous across the infant and the adult (see Figure below).

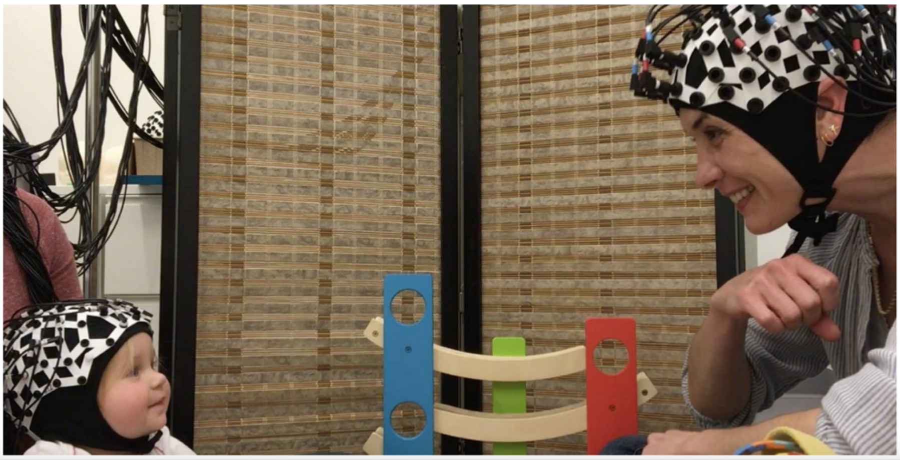


Figure 6. "Example of an interaction between an adult and infant during the together condition (a) and the corresponding intersubject correlation (ISC) in one channel pair (b). The graph shows the concentration of deoxyhemoglobin across the length of an interaction during the together condition, separately for the adult and infant. The ISC between the adult and infant, computed from a single right prefrontal cortex (PFC) channel, is shown at the bottom right" (Piazza *et al*. 2020: 8)

The study reveals the following results. First, as expected, the authors found that time-locked neural coupling within dyads was significantly greater when dyad members interacted with each other than with control individuals. In the together (but not the apart) condition, they found significant coupling between *many prefrontal cortex PFC channels* and *some parietal channels* of infants and the adult, as the figure below shows:

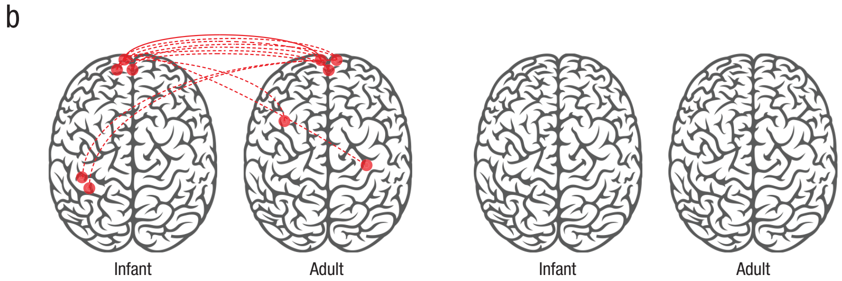


Figure 7. "The brain diagrams (b) show significantly coupled channel pairs (in red), determined by phase-scrambling analysis and corrected for multiple comparisons using the false-discovery-rate procedure (q <.05; Benjamini and Hochberg, 1995). The solid line indicates a homologous channel pair; dotted lines indicate nonhomologous channel pairs." (Piazza *et al.* 2020: 10)

When they looked at the data, the researchers found that during the face-to-face sessions, the babies' brains were synchronized with the adult's brain in several areas known to be involved *in high-level understanding of the world*—rather than perhaps helping the children decode the overall meaning of a story or analyze the motives of the adult reading to them. When the adult and infant were turned away from each other and engaging with other people, the coupling between them disappeared. This result is expected because the same was observed in several other experiments.

7. SNR-IT Superiority

In this final section, we argue that our version of it—what we call 'SNR-IT'—has all the advantages and none of the disadvantages of standard IT. It can better accommodate available data from developmental psychology and neuroscience than its rivals. However, as claimed before, presenting the full range of evidence in developmental psychology is beyond the scope of this paper (see Reddy 2008). Rather, we discuss a few findings related to the notion of social interaction.

First, we consider evidence showing that immediately after birth, neonates imitate some facial gestures of adults (see Meltzoff and Moore 1977). Whether neonates truly imitate or whether their behavior is just a reflex is still a debatable open question. Be that as it may, let us assume just for the sake of argument that there is true imitation going on. The question is: what is the best explanation? According to Gallagher:

> IT treats neonate imitation as a beginning point for what Trevarthen (1979) calls "primary intersubjectivity". Primary intersubjectivity involves early developing sensory-motor processes that already characterize the first year of life. During that first year, infants consistently engage in second-person interactive processes with others, as evidenced by the timing and emotional response of infants' behavior. Infants

> "vocalize and gesture in a way that seems [affectively and dynamically] 'tuned' to the vocalizations and gestures of the other person" (Gopnik and Meltzoff 1996: 131). Developmental studies show the very early appearance of and the importance of this timing and coordination in the intersubjective context. In 'still face' experiments, for example, infants are engaged in a normal face-to-face interaction with an adult for 1 to 2 minutes, followed by the adult assuming a neutral facial expression. This is followed by another normal face-to-face interaction. Infants between 3 and 6 months become visibly discouraged and upset during the still face period when mutual interaction breaks down (Tronick *et al.* 1978.) [See the figure and video below]. The importance of interactive touch has also been demonstrated in the still-person effect (Muir 2002) (Froese and Gallagher 2012: 443).

Figure 8. Tronick's facial experiment. It remains one of the most replicated experiments in developmental psychology. https://www.youtube.com/watch?v=f1Jw0-LExyc

Gallagher is right when he claims that "the infant is not simply a passive spectator trying to figure out what is going on, since the infant is drawn into and actively participates in the process" (Froese and Gallagher 2012: 443). Nevertheless, the crucial questions are as follows. First, how can neonates imitate the adults' facial expressions? What lies behind the neonate's imitation? Second, how can babies coordinate their emotions with the emotions of their caregivers? According to IT, the neonate imitates the adults' facial expressions because the adults' facial expressions are "soaked with mindedness", which the neonate *directly and smartly perceives*. Now, is it reasonable to assume that neonates imitate their caregivers' facial expressions because, as active participants in the process, they *directly and smartly perceive* (at least the most basic seven) emotions in the face of their caregivers?

Even though there are no BBC studies with neonates, Piazza and colleagues' studies with preverbal infants give us a better idea of what is going on in neonates' imitation of their caregivers' facial expressions. First, IT claims that the neonate's imitation is the beginning of "primary intersubjectivity", which supposedly involves early developing sensory-motor processes. However, Piazza and colleagues' studies revealed no significant activation of the PMC and dPMC. Instead, their study indicates that infants' and adults' brains synchronize by higher-order areas such as the PFC and some parietal cortex. Moreover, in both brains, the PFC activation slightly preceded moments of mutual gazing (joint eye contact between the infant and adult). Specifically, the relationship between the brain and behavior peaked at approximately 5 s before the initiation of the mutual gaze. This suggests that the PFC of both individuals anticipated—or even drove—an increase in joint social behavior. Again, these data are unexpected in light of the radical enactivist-based IT. For one thing, according to radical enactivists, embodied behavior constitutes perception. However, the evidence fits with SNR-IT, as the coupling between the neural activities of the caregiver and the infant occurs through their neural situated representations firing in PFC. The sharing of body activity and the environment plays a crucial role in the simulation that these representations make. But what can one say about secondary subjectivity? According to Gallagher:

> Secondary intersubjectivity begins with joint attention during the first year (Trevarthen and Hubley 1978; Reddy 2008). Infants begin to enter into the kinds of interactions with their caregivers that involve jointly attending to objects in pragmatically or socially contextualized situations (Kaye 1982). They learn from others about the value or usefulness of certain objects and actions. They begin to see that another's movements and expressions often depend on meaningful and pragmatic contexts and are mediated by the surrounding world (Froese and Gallagher 2012: 444).

However, Piazza and colleagues revealed that brain-to-brain coupling was greater in eye-to-eye gazing than in joint attention. Moreover, Piazza and colleagues' studies also reveal that the PFC of both individuals anticipated—or even drove—an increase in joint social behavior. Consistent with folk psychology, they hypothesize that the PFC predicts behaviors, contributes to the generation of those behaviors, or responds to earlier social events.

Thus, on the basis of the SNR-IT framework, we hypothesize that when a neonate imitates their caregiver's facial expressions, the neonate's brain and the adult's brain synchronize via the same PFC and some parietal cortex or similar higher-order areas using SNR. To be sure, we are not disregarding the crucial role that the body, action, and environment play; we are simply rejecting a radical version of enactivism that dismisses the role of representations in explaining mindreading. Hence, it is not passive mental representations or motor processes that underpin IT, but neural representations whose contents are necessarily constructed from bodily activity, affectivity, and dynamics in the sharing environment.

8. Concluding Remarks

Let us reconsider the case of the infant and caregiver. As before, they share a neural pattern homomorphic to their target, such as a map of the room where they are playing. Second, the prefrontal cortex (PFC), temporoparietal junction, and parietal cortex (areas involved in prediction, language processing, and understanding others' perspectives) activate in response to the same target, such as the toys. Third, this shared neural pattern controls behavior to achieve their goals.

However, as we discussed, the structural representation of the goal depends crucially on the *ongoing simulation* of both the environment and the bodies involved: the toddler's and caregiver's bodies. The simulation regulates behavior through motor commands and continuously updates with information from the body, environment, and motor commands.

But where does this lead us? It leads to the idea that synchronized cortical areas (the PFC, temporoparietal junction, and parietal cortex) function *as situated neural representations* (SNR). Above all, brainets require embodiment—the dynamic connection between the neurocognitive system and the body. This embodiment is crucial not only because the body houses the sensors and effectors that allow the neurocognitive system to receive and act on information but also because the real-time feedback loop between the neurocognitive system and the body plays a critical role.

The synchronized cortical areas serve as situated neural representations for another reason: *they are embedded within an environment*. This is significant because the environment provides the primary source of information for most senses, and the body depends on it to function. Moreover, the real-time feedback loop between neurocognitive systems and their environment is essential for accurately representing the environment, distinguishing it from the body, and effectively simulating and interacting with it.

The third reason is that the synchronized cortical areas act as situated neural representations due to enaction. Each motor command affects:

(1) Body movement.

(2) Changes in sensory input due to alterations in the body's position relative to the environment.

(3) Changes in the environment as the organism's actions alter it.

Finally, these cortical areas function as SNR, because they involve, affect/motivation. Toddlers and caregivers with behavioral flexibility rely on affect to choose actions that meet their needs, assess external situations, and eventually learn to select adaptive action sequences in different contexts. This becomes clear when the caregiver shifts its attention away from the toddler.

We claim that brainets, understood as shared situated neural representations, help us move away from metaphors that suggest individuals "smartly and directly perceive the mental states of others" or that bodily expressions are inherently "imbued with mindedness". On closer inspection, there is no such "smart perception" of others' mental states, nor are bodily expressions "imbued with mindedness". Instead, specific areas of the brain—the prefrontal cortex, temporoparietal junction, and parietal cortex—function as situated neural representations: embodied, embedded, enactive, and affective representations. So, representations that involve the body, affectivity, action, and the dynamics of the environment play the crucial role in mindreading the minds of others. This constitutes the most fundamental aspect of what it means to be "the sense of us".[13]

---

[13] This work was supported by CNPq [PQ–1B] and CAPES.

References

Baron-Cohen, S., 1995. *Mindblindness: an essay on autism and theory of mind*. Cambridge, MA: The MIT Press.

Berent, I., 2020. *The blind storyteller: how we reason about human nature*. Oxford: Oxford University Press.

Birch, J., 2024. *The edge of sentience: risk and precaution in humans, other animals, and AI*. Oxford: Oxford University Press.

Block, N., 2023. *The border between seeing and thinking.* New York: Oxford University Press.

Carey, S., 2009. *The origins of concepts*. Oxford: Oxford University Press.

Carruthers, P., 1996. Autism as mind-blindness: an elaboration and partial defense. *In*: Carruthers, P., Smith: K. (eds.) *Theories of theories of mind*. Cambridge: Cambridge University Press.

Carruthers, P., 2015. Perceiving mental states. *Consciousness and cognition*, 36, 498–507.

Carruthers, P., 2024. *Human motives: hedonism, altruism, and the science of affect*. New York: Oxford University Press.

Cassam, Q., 2007. *The possibility of knowledge*. Oxford: Clarendon Press.

Chemero, A., 2009. *Radical embodied cognitive science*. Cambridge, MA: MIT Press.

Clark, A., 1997. *Being there: putting brain, body and world together again*. Cambridge, MA: MIT Press.

Clark, A., 2015. *Surfing uncertainty: prediction, action, and the embodied mind.* Oxford: Oxford University Press.

Crone, K., 2020. Foundations of a we-perspective. *Synthese*, 12, 1–18.

Davidson, D., 2001. Rational animals. Reprinted in his *Subjective, intersubjective, objective*. Oxford: Clarendon Press, 95–105.

Dehaene, S., 2014. *Consciousness and the brain: deciphering how the brain codes our thoughts*. New York: Penguin Books.

Dehaene, S., 2020. *How we learn: why brains learn better than any machine…for now*. New York: Viking.

De Waal, F., 2019. *Mama's last hug: animal emotions and what they tell us about ourselves*. New York: Norton and Company.

Di Paolo, E.A., Cuffari, E.E., and De Jaegher, H., 2018. *Linguistic bodies: the continuity between life and language*. Cambridge, MA: MIT Press.

Froese, T. and Gallagher, S., 2012. Getting interaction theory (IT) together: integrating developmental, phenomenological, enactive, and dynamical approaches to social interaction. *Interaction studies*, 13 (3), 436–468.

Gallagher, S., 2004. Understanding interpersonal problems in autism: interaction theory as an alternative to theory of mind. *Philosophy, psychiatry, and psychology*, 1, 199–217.

Gallagher, S., 2008. Direct perception in the intersubjective context. *Consciousness and cognition*, 17 (2), 535–543.

Gallagher, S., 2012. In defense of phenomenological approaches to social cognition: interacting with the critics. *Review of philosophy and psychology*, 3 (2), 187–212.

Gallagher, S. and Hutto, D., 2008. Understanding others through primary interaction and narrative practice. *In*: J. Zlatev, T. Racine, C. Sinha and E. Itkonen (eds.), *The shared mind: perspectives on intersubjectivity*. Amsterdam: John Benjamins.

Green, M., 2010. Perceiving emotions. *Proceedings of the Aristotelian society, Supplementary volume*, 84, 45–61.

Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., and Keysers, C., 2012. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences*, 16 (2), 114–121.

Hutto, D.D. and Myin, E., 2012. *Radicalizing enactivism: basic minds without content*. Cambridge, MA: MIT Press.

Hutto, D.D. and Myin, E., 2017. *Evolving enactivism: basic minds meet content*. Cambridge, MA: MIT Press.

Husserl, E., 1995. *Cartesian meditations: an introduction to phenomenology*. Trans. D. Cairns. Dordrecht: Kluwer.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L. and Lu, C., 2012. Neural synchronization during face-to-face communication. *The journal of neuroscience: the official journal of the society for neuroscience*, 32 (45), 16064–16069.

Jacob, P., 2011. The direct-perception model of empathy: a critique. *Review of philosophy and psychology*, 2 (3), 519–540.

Krueger, J., 2012. Seeing mind in action. *Phenomenology and the cognitive sciences*, 11, 149–173.

McNeill, W.E.S., 2012. Embodiment and the perceptual hypothesis. *The philosophical quarterly*, 62, 569–591.

Meltzoff, A.N. and Moore, M.K., 1977. Imitation of facial and manual gestures by human neonates. *Science*, 198 (4312), 75–78.

Mill, J.S., 1865. *An examination of Sir William Hamilton's philosophy*. 4th ed. London: Longman, Green, Reader and Dyer.

Muir, D. and Lee, K., 2002. The still-face effect: methodological issues and new applications. *Infancy*, 4, 483–491.

Myin, E., 2016. Perception as something we do. *Journal of consciousness studies*, 23 (5–6), 80–104.

Neander, K., 2017. *A mark of the mental: in defense of informational teleosemantics*. Cambridge, MA: MIT Press.

Neupane, S., Fiete, I. and Jazayeri, M., 2024. Mental navigation in the primate entorhinal cortex. *Nature*, 630, 704–711.

Nicolelis, M., 2020. *The true creator of everything: how the human brain shaped the universe as we know it*. New Haven and London: Yale University Press.

Nguyen, T., Abney, D.H., Salamander, D., Bertenthal, B.I. and Hoehl, S., 2021. Proximity and touch are associated with neural however, not physiological synchrony in naturalistic mother-infant interactions. *NeuroImage*, 244.

Nguyen, T., Schleihauf, H., Kayhan, E., Matthes, D., Vrtička, P. and Hoehl, S., 2019. The effects of interaction quality on neural synchrony during mother-child problem solving. *Cortex*, 124, 235–249.

Noë, A., 2005. *Action in perception*. Cambridge, MA: MIT Press.

Oberman, L.M. and Ramachandran, V.S., 2015. The role of the mirror neuron system in the pathophysiology of autism spectrum disorder. *In*: Frances, F., Giacomo, R. *New frontiers in mirror neurons research*. Oxford: Oxford University Press.

Piccinini, G., 2021. *Neurocognitive mechanisms: explaining biological cognition*. Oxford: Oxford University Press.

Piccinini, G., 2022. Situated neural representations: solving the problems of content. *Frontiers in neurorobotics*, 16.

Piccinini, G., forthcoming. Embodied (4EA) cognitive computational neuroscience. *Cognitive neuroscience*.

Ramakrishnan, A., Ifft, P.J., Pais-Vieira, M., Byun, Y.W., Zhuang, K.Z., Lebedev, M.A. and Nicolelis, M.A.L., 2015. Computing arm movements with a monkey brainet. *Scientific reports*, 5.

Ratcliffe, M. J., 2007. *Rethinking commonsense psychology: a critique of folk psychology, theory of mind and simulation*. Basingstoke: Palgrave Macmillan.

Reddy, V., 2008. *How infants know minds*. Cambridge, MA: Harvard University Press.

Ryle, G., 1949. *The concept of mind*. London: Hutchinson.

Spelke, E.S., 2022. *What babies know: core knowledge and composition*. Vol. 1. Oxford: Oxford University Press.

Tronick, E., Als, H., Adamson, L., Wise, S. and Brazelton, T.B., 1978. The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American academy of child psychiatry*, 17 (1), 1–13.

Tseng, P.-H., Rajangam, S., Lehew, G., Lebedev, M.A. and Nicolelis, M.A.L., 2018. Interbrain cortical synchronization encodes multiple aspects of social interactions in monkey pairs. *Scientific reports*, 8.

Watson, J.B., 1913. Psychology as the behaviorist views it. *Psychological review*, 20, 158–177.

Zahavi, D., 2011. Empathy and direct social perception: a phenomenological proposal. *Review of philosophy and psychology*, 2, 541–558.