Mentalizing Homeostasis: The Social Origins of Interoceptive Inference

Aikaterini Fotopoulou[1] & Manos Tsakiris[2]

[1] Clinical, Educational & Health Psychology, Division of Psychology & Language Sciences, University College London, UK

[2] Lab of Action & Body, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, UK

Correspondence concerning this article should be addressed to:
Aikaterini (Katerina) Fotopoulou
CEHP Research Department
University College London
1-19 Torrington Place
London WC1E 7HJ
Tel: (+44) 020 7679 3079
Fax: (+44) 020 7916 8502
Email: a.fotopoulou@ucl.ac.uk.

Keywords

Self; affect; emotion; awareness; social cognition; touch; interoception; intersubjectivity; embodiment; skin ego; minimal self

Acknowledgements
A.K. was supported by a ‘European Research Council Starting Investigator Award' [ERC-2012-STG GA313755]. MT was supported by the European Research Council (ERC-2010-StG-262853) under the FP7, ESRC Research Grant ES/K013378/1 and Bial Foundation Research Project 087/2014.

Abstract

Is the self already relational in its very bodily foundations? The question of whether our mental life is initially and primarily shaped by embodied dimensions of the individual or by interpersonal relations is debated in many fields, including psychology, philosophy, psychoanalysis, and more recently, cognitive neuroscience. In this interdisciplinary target article, we put forward the radical claim that even some of the most minimal aspects of selfhood, namely the feeling qualities associated with being an embodied subject, are fundamentally shaped by embodied interactions with other people in early infancy and beyond. Such embodied interactions allow the developing organism to mentalize its homeostatic regulation. In other words, embodied interactions contribute directly to the building of mental models of the infant's physiological states, given the need to maintain such states within a given dynamic range despite internal or external perturbations. Specifically, our position rests on the following three propositions: (1) The progressive integration and organisation of sensory and motor signals constitutes the foundations of the minimal self, a process which we have linked to contemporary, computational models of brain function and named 'embodied mentalization'; (2) Interactions with other people are motivated and constrained by the same principles that govern the 'mentalization' of sensorimotor signals in the individual --and hence the mentalization of one's body can include signals from other bodies in physical proximity and interaction, especially in interaction with *particular* bodies. (3) Crucially, given the dependency of humans in early infancy, there is a 'homeostatically-necessary' plethora of such embodied 'proximal' interactions, especially as regards interoception. Collectively, such experiences of proximal intercorporeality 'sculpt' the mentalization process and hence the constitution of the minimal self, including the progressive sophistication of mental distinctions between 'subject-object', 'self-other' and even 'pleasure-pain'. Finally, we explore notions of cardiac and more broadly

interoceptive awareness, as later, cognitive acquisitions that allow us to progressively solidify such distinctions, as well as understand and empathise with other people.

1. Introduction: Is the Self Primarily Bodily or Social?

A child falls in the playground and mildly scrapes her knee. What is her first reaction? Begin to cry? Look at the knee to verify the external location and degree of the tissue damage she presumably feels from the inside? Both are possible. Yet, as many parents would attest to, frequently the first thing that children do is turn to the parent and await their reaction before they proceed with a proportional, behavioural reaction of their own. Developmental and social psychologists explain such behaviors as related to modelling (Bandura, 1967) and social referencing (Klinnert et al., 1986). Psychodynamic scholars may interpret such behaviors as identification (Hobson & Lee, 1999), 'affect attunement' (Stern, 1985), or affect regulation in the context of attachment (Bowlby, 1969, 1973). However, such behaviors raise a more fundamental question. Why does an experience of bodily pain, so intimately connected with subjectivity and an individual's own body, invite immediate social attention and reaction? In this interdisciplinary target article, we put forward the radical claim that even some of the most minimal aspects of selfhood, namely the feeling qualities associated with being an embodied subject, are fundamentally shaped by embodied interactions with other people in early infancy and beyond.

Many fields have examined whether our mental life is initially and primarily shaped by embodied dimensions of the singular individual or by interpersonal relations. In psychology, philosophy, and psychoanalysis, several debates surround the question of whether bodily or social drives are the primary motivator of the mind, as well as the question of whether bodily or social experience is the primary organizer of the self (see also Ciaunica

& Fotopoulou, 2016). In psychoanalysis, for example, Freud sketched the development of the self as dependent mostly on the vicissitudes of unconscious and initially 'objectless' drives. In addition, he stressed the bodily, and predominately sensory, origins of selfhood. Object-relations theory, by contrast, articulated by psychoanalysts such as Ronald Fairbairn and Melanie Klein, emphasized that the primary motivation in infancy is object seeking (social relating) and the primary organizer of the mind are internalized social relations. Attachment theory, as developed by John Bowlby (1969, 1973, 1980), expanded by developmental researchers such as Mary Ainsworth (et al., 1978), and integrated into developmental, psychodynamic theory more generally (e.g. Fonagy, Gergely, Jurist & Target, 2002), further contributed to the idea of a primary social drive and a socially constructed self. While contemporary psychoanalytic models have developed beyond these classic theories and several intermediate positions have been put forward (e.g. Winnicott, 1972; Stern, 1985; Anzieu, 1989; Beebe & Lachmann 1998; Aron & Sommer Anderson, 1998), fundamental disparities still exist between theories that emphasize the individual versus the social origins of the mind.

In addition, one can trace a parallel 'de-somatization' tendency within psychoanalysis. Theoretical developments have progressively abandoned not only the centrality of sexuality and other bodily drives in mental life (e.g. see Aron, 1998; Green, 1997; Fonagy, 2008 for critical discussions), but most perspectives within psychoanalysis have paid progressively less attention to the once core idea that the mind, and particularly the self, are rooted in and structured by embodied and enacted experiences. As Conger has reflected, "the body has been invisible, for years unaddressed and ignored, left in the waiting room of the therapist's office" (Conger, 1994, pp. 211). Instead, several analysts have called for emphasis on the symbolic, metaphorical, 'meaning-based' and more generally psychological, or mentalistic representation of bodily phenomena (e.g. Greenberg, 1991; Gill, 1994; Aron, 1998). This trend can be seen even in the relation of psychoanalysis and attachment theory. For example, Bowlby's theories emphasize both embodied concepts such as 'physical proximity' to the caregiver, and more cognitive notions such as 'internal working models' (Bretherton and Munholland, 1999), as key developmental ingredients of psychological growth. Yet attachment research and theory have since focused increasingly on mentalistic concepts such as 'maternal sensitivity' (Cassidy & Shaver, 2008), the regulation of emotional states (Slade 2000), and eventually metacognitive concepts such as 'mentalizing' and 'reflective functioning', i.e. the ability to infer and understand the mental states of oneself and others (Fonagy, Gergely, Jurist, & Target, 2002; see also Fonagy & Target, 2007 and Shia & Fonagy, 2016 for critical discussions). It seems that in some respects, contemporary developments in psychoanalytic theory have prioritized the symbolic, reflective and narrative aspects of the self at the expense of its embodied nature (e.g. see Fonagy & Target, 2007; Conger, 1994 for critical reviews).

Similar debates and developments can be traced in philosophy, in developmental psychology and more recently in cognitive neurosciences. In philosophy, for example, Zahavi (2014, 2015a,b) has recently presented a careful analysis of two influential, contrasting positions on the nature of the self. On the one hand, (embodied) experiential minimalism claims that our mental life is characterized by a pre-reflective sense of self or "mineness" which can be traced to the body, and can and should be understood without any contrasting *others*. In this view, we initially experience ourselves as a "self" because of our own embodied experience. On the other hand, according to social constructivist views, the self is not innate, but a later socio-culturally determined acquisition, emerging in the process of social exchanges and narrative practices (e.g. Schechtman, 1996). Thus, this view argues that there can be no sense of self without an engagement with others.

In developmental psychology, the question of the embodied versus the social origins of the self has centered on whether there is a rudimentary distinction between self and other in infancy (e.g. Gallagher & Meltzoff, 1996), or whether the infant and the caregiver can be best considered as an undifferentiated system from the point of view of the infant's mind (e.g. Welsh, 2006). Several different views have been put forward on each side of this debate. Views that do not recognize an early self-other distinction include, for example, Piaget's classic 'cognitive' views on the fused but egocentric infant, or Winnicott's more affective, holding, mirroring and undifferentiated mother-infant dyad (see Müller et al., 2006 for review). By contrast, proposals for an early self-other distinction include Meltzoff's imitating infant (1989) that can differentiate between self and other from infancy, and Daniel Stern's (1985) clearly differentiated but 'affectively attuned' infant.

Furthermore, these two positions on the nature of the self-other distinction have implications for the understanding of the development of social cognition and self-other relatedness. For scholars that view the infant's mind as socially undifferentiated (e.g. Merleau-Ponty, 1960), the question of 'other minds' is relatively straightforward, in the sense that the other's mental states are not opaque, they do not require any reflective inference but are instead directly perceivable in their behavioral expressions. In the words of Scheler, it is "*in* the blush that we perceive shame, *in* the laughter joy" (Scheler, 1913/70, p. 10, emphasis as in original). Thus, for these scholars the critical developmental task is one of separation, differentiation and individuation (Mahler et al., 1973). By contrast, scholars that adhere to the possibility of an early, rudimentary distinction between self and other are mostly interested in describing how the understanding of other minds can ever be possible, as minds are seen as private and directly 'unperceivable' (Meltzoff & Moore, 2000; Morton & Frith, 1995; Gallese, 2005; Tomasello et al., 2005). This latter question has centered on the developmental trajectory of the so-called ‘theory of mind’, or ‘mentalization’ capacity, defined as the ability to understand and attribute mental states to oneself and others.

Interestingly, empirical data have not really settled these debates for or against the ‘theory of mind’ accounts. For example, according to one view, young infants, as well as non-human primates, are able to ‘mind-read’ to some degree, while according to other views, such abilities in early infancy are best characterized as ‘behavior reading’, or ‘goal understanding’ (Poulin-Dubois et al., 2009). Hence, they should not be viewed as evidence in favor of the existence of early ‘theory of mind’ abilities. According to the latter interpretations, infants may ‘pass’ theory of mind tests because they are able to estimate the statistical likelihood that some behaviors (e.g., gaze direction) will be linked to future actions (e.g., reach towards a congruent location). Thus, distinguishing between mentalistic and expectancy violation accounts has proven rather difficult (Povinelli & Vonk, 2004). Similarly, the results of infant imitation experiments have generated more interpretations and debates than the questions they set up to answer (e.g. compare Gallagher and Meltzoff, 1996 with Welsh, 2006).

Moreover, recent decades show intense debates even among the proponents of the idea that an early distinction between self and other necessitates the development of ‘mentalization’ abilities. For example, scholars disagree on whether understanding other minds is achieved by ‘simulation’ and ‘analogy’ with one’s own, first-person, embodied perspective (e.g. Gallese, 2005), or by cognitive inference from a third-person perspective (e.g. Morton & Frith, 1995). Indeed, more recently a number of intermediate positions have been put forward, notably interaction theory (Fuchs & De Jaegher, 2009; Gallagher, 2001, 2004) and second-person approaches (Reddy, 2008; Schilbach et al., 2013). These emphasize neither the first- nor the third-person view on understanding other minds, but instead put forward the idea that early, reciprocal interactions and emotional engagements with caregivers are fundamental to build shared 'we-experiences' that progressively form the basis of all social understanding. There are however on-going debates regarding the precise role of the so-called "we-experiences" or "plural self-awareness" and whether these we-experiences presuppose, or lead to the self-other differentiation (Reddy, 2008; Gallotti and Frith, 2013; Schmid 2014; Zahavi & Rochat, 2015).

The self is both bodily and social: A summary of our main claim and its implications

Our position is clearly in line with the above second-person and interaction theories (see also Ciaunica & Fotopoulou, 2016). Nevertheless, as we explain in detail below, we put forward an alternative second-person proposal that places emphasis on the fact that one of the main purposes of early social interactions is the regulation of the infant's homeostasis. Thus, our aim is to present a rather reductionistic and mechanistic account of how embodied interactions lead to the constitution of minimal, affective selfhood in development and beyond. To explain our position, we will start by describing briefly what we mean by a 'minimal affective selfhood'. Nevertheless, it should be clear by now that this target article cannot possibly address in full the many questions and debates on the nature of the self, within and across fields, nor are we aiming for a comprehensive review of the relevant empirical literature on the self-other distinction and the understanding of other minds. Instead, we aim to merely show that some of the aforementioned divisions and debates are at least partly fueled by the tacit over-reliance on the role of visual, proprioceptive and verbal modalities in the constitution of the self at the expense of other modalities such as affective touch, pain and interoception (in its wider, homeostatic definition, see below). We argue that when one reconsiders the role of these interoceptive modalities in infancy and beyond, the bodily and the social origins of the self appear as tightly interwoven. In a similar vein, the assumption of a sharp distinction between behavior- and mind-reading becomes less meaningful. Indeed, we will argue that inferential processes such as 'mentalization' or 'theory of mind' are actually advanced forms of more primitive inferential processes of embodied perception and action, what we will term 'embodied mentalization'.

Moreover, this progressive 'mentalization' of the body in development includes not only the body of the singular individual but other bodies in proximity and interaction. Thus, building a mind, and understanding other minds, are embodied and tightly connected processes. We argue that the constitution of the self is dependent upon the social mentalization of the body and particularly its homeostatic needs. In short, the radical aspect of our proposal is that social interactions do not shape *only* the reflective (narrative or extended) self and related notions of affect regulation and social cognition. Instead, *the most minimal aspects of selfhood*, namely the feeling of being an embodied, agentive subject, *are fundamentally shaped by embodied interactions with other people* in early infancy and beyond. Progressively these embodied interactions allow the developing organism to mentalize its homeostasis and hence they constitute the core of our embodied subjectivity. We unpack these ideas below.

2. The Minimal Self: Pre-reflective, "Ego-logical" and Affective

The question of what, if anything, provides a sense of self or makes the "self" a unifying phenomenon has attracted a considerable number of philosophical enquiries and empirical studies, the review of which lies beyond the scope of the present paper. Rather, for our limited purposes we note that, despite disagreements on crucial questions about the existence and nature of the self, several contemporary accounts share the assumption that selfhood is not a subjective reflection on some other mysterious substance, or structure called the "self" (see also Ciaunica & Fotopoulou, 2016). Instead, both classic phenomenologists such as Husserl and Merleau-Ponty, and more recent scholars working within the embodied/enactive cognition paradigm (Varela et al. 1991; Gallagher 2000; Zahavi 2005), insist on the idea that the foundations of our self-awareness are bodily. According to Merleau-Ponty (1945) for example, the basis of our self-awareness does not involve any reflective, objectifying of the self, but is instead experienced as a lived sense of practical engagement and orientation. It is not reflexive consciousness (ideas, beliefs, knowledge) that constitute the world, but a body which is already of and in this world. Our embodiment brings to experience an a priori structure, tacitly expressed in our possibilities for action. From this perspective, several contemporary philosophers (Legrand, 2006; Zahavi, 2014) insist that our minimal self is *pre-reflective*, in the sense that we can be aware of our embodiment without needing any kind of self-oriented thought that presupposes an epistemic division between subject and object.

Moreover, some phenomenological accounts suggest that pre-reflective awareness is a fundamentally *first-personal* subjective experience (Zahavi, 2006); it possesses content that concerns oneself ("egological"). While not all philosophers would agree with this suggestion (e.g. Janzen, 2007; Krueger, 2011), for some philosophers (Zahavi, 2006), this first-personal content is synonymous with the feeling qualities that are intrinsic to phenomenal consciousness, the 'what-it-is-like' qualities of subjectivity – the 'what-it-is-like' feeling of seeing or eating an apple, for example. This description does not imply a cognitive act of attributing such feelings and sensations to a distinct entity (e.g. the self) from the object of consciousness (e.g. a perceived apple). Rather this description implies a fundamental property, namely the direct giveness of such feelings and sensations as *my* feelings and sensations. Their existence as experiential states constitutes subjectivity in the same breath as subjectivity constitutes their existence as experiential states. As Zahavi describes it, "A conscious mental state is not merely conscious of something, its object, it is simultaneously self-disclosing or self-revealing" (Zahavi, 2016).

Interestingly, in his recent influential neuropsychoanalytic account of self-consciousness, Solms (2013) seems to put forward a similar view regarding the *affective* core of subjectivity, arguing in favor of an embodied, affective consciousness which forms the background of all subjective, conscious experience. Solms draws a sharp distinction between an 'inner mental body', monitoring homeostatic needs and manifesting as affective consciousness along a pleasure-unpleasure dimension, and perceptual consciousness, which he links to the classic senses and the perception of the 'outer body', animated only via its connection to the inner body. In his words,

> *"The internal type of consciousness consists in states rather than objects of consciousness (cf. Mesulam 2000). The internal body is not an object of perception unless it is externalized and presented to the classical senses; it is the subject of perception. It is the background state of being conscious. This is of paramount importance. We may picture this aspect of consciousness as the page upon which external perceptions are inscribed…It has recently been recognized that the state of the body-as-subject involves not only varying levels of consciousness (e.g. sleep/waking) but also varying qualities of consciousness (Panksepp 1998, Damasio 2010). The internal aspect of consciousness 'feels like' something. Above all, the phenomenal states of the body-as-subject are experienced affectively. Affects do not emanate from the external sense modalities. They are states of the subject. These states are thought to represent the biological value of changing internal conditions (e.g. hunger, sexual arousal). When internal conditions favour survival and reproductive success they feel 'good', when not they feel 'bad'. This is evidently what conscious states are for. Conscious feelings tell the subject how well it is doing. At this level of the brain, therefore, consciousness is closely tied to homeostasis". (Solms, 2013 p. 7).*

Clearly, there is much more to be said about this neuropsychoanalytic perspective on self-consciousness, the above phenomenological insights, and the related debates in contemporary philosophy of mind. However, for the purposes of this paper we will restrict ourselves to clarifying our central claim: namely, while we broadly agree with the above neuropsychoanalytic and phenomenological views regarding the embodied and affective origins of subjectivity and selfhood, we are suggesting that at least certain parts of our embodied, affective subjectivity are *interpersonally constituted*. The important role of other people in the formation of the self is acknowledged by both accounts. However, it is regarded as relevant to later, cognitive acquisitions of the extended, or narrative selfhood, rather than being constitutive of the affective core of subjectivity and minimal selfhood.

Moreover, we should highlight that by arguing that the phenomenal quality of conscious states is interpersonally constituted we do not mean to imply that infants without caregivers would not have an affective minimal self at all and they would be in some unconscious ‘zombie-like’ state (see Zahavi, 2016 for criticism and the description of a ‘thinner’ minimal self than the one discussed here). We believe that the capacity for a minimal, affective consciousness is prescribed by phylogenetic development, but nevertheless each infant’s minimal self (i.e. the particular quality of its experiential states) is determined in ontogenetic development. The evolutionary risk of lacking caregivers is not some unconscious ‘zombie-like’ state, but rather death. Hence, under the assumption that the mind was developed to serve the survival needs of the body, it is plausible to assume that one of the key features of the mind, namely its affective, experiential core, was developed to serve the survival needs of infants’ bodies in relation to the very people that can ensure such survival in ontogenesis. Against this theoretical background, we aim to argue in favor of a reconceptualization of minimal selfhood that traces the relational origins of the self on fundamental principles and regularities of the human embodied condition, which necessarily includes social, embodied interactions and practices in early development and beyond (see also Ciaunica & Fotopoulou, 2016).

Of course, this article cannot do justice to all the possible philosophical implications of this position but rather based on current neuroscientific knowledge we are putting forward a proposal regarding the potential mechanisms that allow the minimal self to be interpersonally constituted. Specifically, our position is motivated by the following three theoretical and empirical observations:

(1) The progressive integration and organization of sensory and motor signals constitutes the foundations of the minimal self, a process which we have linked to contemporary, computational models of brain function and named ‘mentalization’ elsewhere (Fotopoulou, 2015);

(2) Interactions with other people are motivated and constrained by the same principles that govern the ’mentalization’ of sensorimotor signals in the singular individual and hence the mentalization of one’s body can include other bodies in physical proximity and interaction; and

(3) Crucially, given the premature birth, the immature motor system, and social dependency of humans in early infancy, there is a ‘homeostatically-necessary’ plethora of such embodied ‘proximal’ interactions, particularly as regards interoception and particularly with *some* bodies.

Collectively, such experiences of proximal intercorporeality ‘sculpt’ the mentalization process and hence the constitution of the minimal self, including the progressive sophistication of mental distinctions between ‘subject-object’, ‘self-other’, and even ‘pleasure-pain’. We unpack these points below, by focusing predominately on the domain of social touch, as a paradigmatic example of proximal intersubjectivity, and affective touch, as a paradigmatic example of shared interoception. We subsequently explore notions of cardiac and more broadly interoceptive awareness, as later, cognitive acquisitions. Interoceptive awareness allows us to progressively solidify the boundaries of the self and psychologically ‘separate’ the self from the other, as well as ultimately ‘mentalize’ both as objects of our perception.

3. Embodied Mentalization: A Free Energy Perspective

Beyond philosophical debates, and not always in agreement with all the above philosophical subtleties, the emphasis on a minimal, embodied self is adopted also by recent influential models of brain function in theoretical neuroscience (Friston, 2010; see also Clark, 2013), as we will consider in greater detail below. In scientific accounts, minimal selfhood can thus be conceived as a dynamic, ongoing process of tracking and controlling bodily properties (Blanke and Metzinger, 2009). This idea has been the focus of much recent, empirical research including investigations that use experimental 'tricks' to systematically manipulate sensorimotor and multisensory signals, promote their integration, or generate conflicts and illusions, and hence study their role in action, perception and body awareness (Tsakiris, 2010; Blanke et al., 2015 for reviews). These studies, as well as investigations in neuropsychiatry (see Fotopoulou, 2012a; Jenkinson & Fotopoulou, 2014; Fletcher & Fotopoulou, 2016 for reviews) suggest that primary multisensory and sensorimotor signals are integrated and organized at different levels of the neurocognitive hierarchy to form several neurocognitively distinct dimensions of minimal, as well as 'extended' selfhood (Farmer & Tsakiris, 2012; Fotopoulou, 2014; 2015). Related notions, for example, are the concepts of 'body ownership' (the pre-reflective sense or metacognitive judgement that bodily sensations and movements and the body parts upon which these are experienced are unique to one's self) and of 'bodily agency' (the pre-reflective sense or metacognitive judgement that I am the cause of a movement and its consequences) (Gallagher, 2000; Legrand, 2006). Unfortunately, due to space restrictions, we are not able to cover the role of motor agency in the current article and we have mainly focused on body ownership.

Recently, we have independently used an influential theory from computational neuroscience, namely the 'Free Energy Principle' (FEP, Friston, 2010) to describe how the integration of sensorimotor signals can be described as the ‘mentalization’ of sensorimotor signals (Fotopoulou, 2015; Besharati et al., 2016), and to propose a model of the multisensory processes that give rise to a facet of the minimal self, namely body-ownership (Apps & Tsakiris, 2014). These two theoretical proposals on the minimal self are important precursors to understanding the more radical, developmental proposal put forward in the current target article. We therefore, first outline the FEP framework (Friston, 2010) and subsequently these two theoretical proposals in turn below (see Carhart-Harris & Friston, 2010; Fotopoulou, 2012b; Solms, 2013; Fotopoulou, 2013 for the relation between this framework and the Freudian metapsychology more generally).

The starting point of the ‘FEP’ framework (Friston, 2005) is that humans are biological, self-organizing agents that need to occupy a limited repertoire of sensory states for homeostatic reasons (for example, humans need to stay within certain ranges in environmental temperature in order to survive). However, due to the inherent ambiguity, complexity and uncertainty of the signals an organism receives from the world, we risk finding ourselves in states for longer periods than those we could biologically sustain (e.g. in cold climates). We thus need to be able to predict (infer) the causes of our possible sensory states despite the limited or noisy information available to our sensory organs (von Helmholtz, 1878/1971). The framework proposes that our brain engages in a form of probabilistic representation of the causes (e.g. the weather) of our future states (e.g. our temperature) on the basis of noisy sensory data; in other terms, it maintains hypotheses (‘‘generative models’’) of the hidden causes of sensory input. Furthermore, it uses such input to constantly update its models, so as to reduce its representational errors over time and thus ultimately minimize the risk of ‘surprise’ (unpredictability, see below for a formal definition). It is precisely these embodied and inferential processes that one of us (AF) has proposed should be understood from the point of view of psychology as *‘embodied mentalization’*

(Fotopoulou, 2015). Although the term ‘mentalization’ is traditionally used in psychology to refer to our cognitive ability to infer the mental states of others and our own’, its alternative use in this context is deliberate. We purposefully use the term “mentalization” to ground this traditional concept in its embodied origin, and highlight that self-awareness is not some ‘add-on’ inferential process of ‘mind-reading’, but rather a more fundamental process of organization and schematization of bodily signals; as such it also directly and necessarily extends to the mentalization of *any*-body (see below).

While a full description of the free energy framework, and its mathematical and neurobiological implementation, goes beyond the scope of the current target article, consistently with the aforementioned work (Fotopoulou, 2015; Besharati et al., 2016), we define ‘embodied mentalization’ here as the inferential brain process by which primary sensorimotor and multisensory signals are progressively integrated and schematized to form multiple, predictive models of our embodied states in given environments. These models are not understood as static body representations in the brain (e.g. ‘body schema’ vs. ‘body image’) but rather as ‘hypothetical’ (probabilistic, inferential), dynamic and generative processes (they are constantly updated against received error signals). Thus, in simple terms, the brain is taxed by evolution to engage in constant mental ‘modelling’ of its environment and its own bodily states. For example, our sensory organs do not just passively perceive the environmental temperature, and then our brain forms mental conclusions about the weather conditions. Instead, our brain anticipates that certain weather conditions will influence our sensory organs and body temperature in a particular way, and we update our expectations about the weather based on the difference between the expected and the experienced body temperature. Thus, contrary to common sense views, our body is not there to serve our mind’s thoughts and desires, but rather the mind is there to serve the body’s needs in a given environment.

More specifically, the free energy framework is biologically constrained by the so-called ‘predictive coding’ models of perception (e.g. Henson & Gagnepain, 2010; McNally et al., 2011). According to such models, a constant filtering of sensations by top-down predictions and a parallel updating of the latter based on prediction errors (signals representing the mismatch between predictions and sensations), with the ultimate goal of minimizing prediction errors, is an imperfect but highly efficient means of perceiving sensations (Rao and Ballard, 1999). This model assumes that our brains achieve the minimization of prediction errors by recurrent message passing among hierarchical level of cortical systems, so that various neural subsystems at different hierarchical levels minimize uncertainty about incoming information by generating a prediction (or a prior belief, see below) and responding to errors (mismatches) in the accuracy of the prediction, or prediction errors. Such prediction errors are passed forward to drive the units in the level above, which encode conditional expectations that optimize top-down predictions to explain away (reduce, inhibit) prediction error in the level below, until conditional expectations are optimized. Such message passing is considered neurobiologically plausible on the basis of functional asymmetries in cortical hierarchies; prediction errors are thought to be conveyed via feedforward connections from lower to higher levels in order to optimize representations in the latter. Predictions from higher-levels are also transferred via feedback connections that have both driving and modulatory characteristics and can suppress prediction errors in lower levels. This hierarchy is thus reciprocal but asymmetric and models the nonlinear generation of sensory input (Adams et al., 2013). Based on such hierarchical, perceptual schemes, the free energy principle rests upon the idea that the brain as a whole works as an Helmholtzian inference machine that is trying to optimize its own model of the world by actively predicting the causes of its sensory inputs (Friston, 2005). Moreover, in this framework, perception is tightly linked to action, which is defined as the inverse way to reduce free energy.

Specifically, while perception reduces free energy by prediction updating and cancelling out prediction errors, action can reduce free energy by changing sensory input. After all, merely representing the world (perceptual inference) cannot take us far in terms of our ultimate goal – surviving in an uncertain world (see above). Psychologically speaking, we may become better in predicting the changes in the environment that act to produce sensory impressions on us, but we cannot on this basis change the sensations themselves and hence ultimately their surprise. By contrast, by acting upon the world we can change its states and therefore 're-sample' the world to ensure we satisfy our predictions about the sensory input we expect to receive. This selective sampling of sensory inputs based on expectations adds accuracy to our predictions about sensory states. This is known as 'active inference'. Active inference can reduce free energy by changing sensory input, while perceptual inference reduces free-energy by changing predictions. If perception is a system for predicting reality, action is one for testing such predictions. In this respect, the framework is consistent with theories of embodied cognition and enactive perception (see Clark, 2013 for discussion) that stress the close link between action and perception. To return to our simplified weather example, if our brain anticipates that certain weather conditions will influence our body temperature in a particular way, one way to reduce predictions errors regarding our body temperature is to walk until we reach the predicted body temperature.

At this point it is important to also acknowledge some of the mathematical components of the model, although as aforementioned it is not possible to do them justice in this interdisciplinary article, nor are we able to fully explain them for the non-specialized reader. We thus offer them here for pure reference and they can be skipped by the non-interested reader. Computational neuroscientists understand the above inferential, predictive process in Bayesian terms (Bayes' theorem describes an optimal procedure for updating the probabilities assigned to a hypothesis in the light of new evidence), in the sense that it relies on a combination of prior beliefs (probability distributions over some unknown cause excluding any sensory data) and new sensory data to update prior beliefs and generate posterior beliefs (probability distributions over some unknown cause after data have been received). Furthermore, in the free energy principle this hierarchical minimization of prediction errors is understood as a minimization of free-energy on the basis of the formal (i.e. mathematical) definition of the latter; a quantity from informational theory that bounds (is greater than) the evidence for a model of data (Hinton & van Camp, 1993). In this case the data is sensory and free energy bounds the negative log-evidence (surprise) inherent in sensory data, under a model of how the data were caused (see Friston, 2010 for the mathematical details). Given some mathematical assumptions, free energy can be thought of as the amount of prediction error in any given level of the system. Minimizing free energy then corresponds to explaining away prediction errors following the principles of Bayes (Friston, 2010). To return to our simple weather example, these mathematical, probabilistic notions are used to formulize how we update our expectations about the weather or fulfill them by action optimally, in light of new sensory evidence such as changes in our body temperature.

In the context of this framework, 'embodied mentalization' is therefore the on-going, dynamic process of maintaining and updating generative models of the likely causes of sensory data from inside the body itself (see also section on Interoception below) and the external world. Progressively, as a consequence of learning, predictions become more encompassing and abstract, so that higher levels of cortical representation in the brain become able to integrate information from multiple lower levels and efficiently predict the causes of multisensory and sensorimotor signals. In this way, multicomponent embodied experiences are ultimately and hierarchically organized into coherent mental categories. When activated, these are consciously experienced as independent from the embodied experiences that gave rise to them (as ‘things’ in themselves), and can in fact be brought to consciousness in a top-down fashion, i.e. irrespective of any on-line sensory experience. For example, the sight of a lemon is tightly and almost automatically associated with its distinctive sour taste, even when we are not tasting it, as though the taste resides in the lemon and not in the contact with our mouth’s taste buds. Furthermore, hearing or reading the word ‘lemon’ immediately brings to mind such properties, i.e. its yellow color, its shape, and all other effects a lemon typically has on our body, including some more implicit, or pre-reflective aspects, such as how we usually grasp it, how it can be perceived and manipulated from different perspectives, etc.

The same principles of progressive schematization and predictive organization apply to our own body, so that the various embodied experiences are organized (bound together) in multisensory and sensorimotor ‘wholes’ that allow us to predict the probable causes of experience of the various parts and functions of the physical body itself. Typically, for example, when we see ourselves in the mirror we have a unified and coherent experience of our body as standing in front of the mirror where we ‘feel’ it to be and not in the mirror space where it is reflected. In this sense, the physical body itself is not passively perceived but actively metalized. Thus, for example, by the time we are adults such ‘hypothetical’ (probabilistic, inferential) and generative models (they are constantly updated against received error signals) allow us to experience a continuity in our embodied experience despite potential minor changes in weight, appearance or form, and we even have a fantasy of how our body is perceived by other people or other visuospatial perspectives (mentalization of the body, Besharati et al., 2016). Thus, in brief, not only is the mentalization of the world and our experience in it tightly linked with our embodiment (*embodied mentalization*, Fotopoulou, 2015), but the experience of our physical body itself is progressively organized and schematized in hierarchical, predictive and generative models (*body mentalization*, Besharati et al., 2016).

In terms of bodily self-awareness, more specifically, Apps and Tsakiris (2014) have used the Predictive Coding framework, described earlier, to account for at least some of the basic processes that govern the multisensory basis of bodily self-awareness, and extended it to account for its malleability. They argued that one's body is processed in a probabilistic manner as the most likely to be "me" (see Limanowski & Blankenburg, 2015 for empirical support). Such probabilistic representations are created through the integration of top-down 'predictions' about the body and of bottom-up "prediction errors" from unimodal sensory systems that are then explained away. The mental representation of the physical properties of one's self are therefore also probabilistic. That is, one's own body is the one which has the highest probability of being "me," as other objects are probabilistically less likely to evoke the same sensory inputs. Interestingly, the empirical evidence from bodily illusions that probe the mechanisms of body-awareness suggest that multisensory integration – especially in the absence of motor signals – functions as a potent cue for constructing body-awareness as well as for altering the boundaries between self and other. To illustrate, in the Rubber Hand Illusion (RHI), one of the most influential experimental models of embodiment, watching a rubber hand being stroked synchronously with one's own unseen hand being stroked causes the rubber hand to be experienced as part of one's body (Botvinick & Cohen, 1998; Tsakiris, 2010). In the RHI, visually observed touch on the skin that is temporally congruent with touch detected by the somatosensory system will become associated with each other, resulting in a prediction of a somatosensory event when contact to the skin is about to occur. In contrast, touch between two other non-corporeal objects will never evoke a somatosensory event, and thus the prior probability of a somatosensory event following touch on such objects is very low. So one's own body is probabilistically likely to become and be the object that touch is predicted to be experienced upon. The visual properties of different body parts will also be perceptually learned such that when any object approaches the body, a somatosensory event will be predicted. Thus, perceptual learning within the free-energy and predictive coding frameworks leads to generative models where aspects of one's body are processed as probabilistically the most likely object (or collection of objects) that when touched, moved, threatened, or acted upon in any way, evokes events in the other sensory systems that detect the state of the body. In short, the notion that there is a "self" is the most parsimonious and accurate explanation for sensory inputs.

One final aspect of the free energy framework needs to be emphasized to clarify the nature of the minimal self from this perspective. In the free energy framework, the challenge of the organism is to navigate the world by sustaining a set of prior beliefs, sufficiently robust that can guide action towards the world despite its changes. At the same time, our generative models of the world must not be so immutable that our responses become fixed, stereotypical and insensitive to unpredicted change. Indeed, an intrinsic component of the free energy framework is that our generative models need to maintain a (Bayes optimal) dynamic balance between their robustness and flexibility. In other words, the organism needs to know when to pay attention to bottom-up prediction errors and hence update its models, or act to satisfy its predictions, and when to ignore and suppress such errors by top-down predictions at higher levels. This balance between bottom-up signals and top-down expectations is achieved by minimizing our uncertainty (optimizing certainty) about prediction errors (between and within sensorimotor modalities). In psychological terms, the organism needs to optimize not only how it interprets sensory data, but also how it processes their reliability and salience in a given context. Depending on how much 'weight' can be attributed to certain priors or sensory data, the organism can control the significance it attributes to the sensory data it uses to update its predictions, or explain away prediction errors. For example, one may not notice mild fluctuations in body temperature as environmental temperature changes on a typical working day, but if such fluctuations were to occur while someone is visiting a foreign country for the first time, they may be more sensitive and responsive to such changes.

In Bayesian terms, organisms do not just need to probabilistically infer the states of the world (*content*: mathematically this can be thought of as the center of a probability distribution), they also need to infer the uncertainty of such states (*context*: the dispersion of such distribution). Sensory signals may be more or less relevant depending on the context in which they are encountered. Hence, optimal inference requires optimizing the *precision* (mathematically inverse dispersion or variance, and hence the inverse of uncertainty) of sensory signals (Feldman & Friston, 2010; Friston et al., 2012). This is important, especially in hierarchical schemes, where precision controls the relative influence of bottom-up prediction errors and top-down predictions. Previous neuroscientific studies have linked this notion of 'precision' with neuromodulation of synaptic gain (such as the role dopamine and acetylcholine may have on synaptic gain, Friston, 2010). As regards sensory data deriving from the world (exteroception), this salience optimization can be seen as flexible attention in perceptual inference (Feldman & Friston, 2010), and as affordance sensitivity (latent action possibilities of cues in the environment) in active inference (Friston et al., 2012). As regards sensations about the state of one's own body (interoception, see also below), optimizing the precision of sensory prediction errors can be seen as increased *interoceptive sensitivity* (see Fotopoulou, 2013; Ainley et al., 2016) and as increased *seeking behaviors* in active inference (see Fotopoulou, 2013; Pezzulo et al., 2015). Moreover, precision operates both within and between modalities. At each level of the hierarchy and taking account of the given context, the brain weighs the relative precision of prediction errors (potentially from different modalities) that inform or revise (potentially multimodal) expectations at higher levels of the hierarchy.

The notion of ‘precision optimization’ turns out to be highly relevant for the coherence of our body awareness, despite the variable, ambiguous and conflicting sensory data available to our body. For example, it has been recently shown that the RHI, the aforementioned test of the feeling of body ownership, as well as the so-called ‘sensory attenuation’ phenomena associated with feelings of motor agency can be accounted for by precision optimization (Brown et al., 2013; Zeller et al., 2014). In the RHI for instance, the brain attenuates the precision of ascending, proprioceptive prediction errors about the actual position of the participant’s own arm (Zeller et al., 2014) in order to accept the more plausible (even if illusory) perceptual hypothesis that it is one’s own body that receives synchronous tactile and visual information, rather than the alternative hypothesis that another body evokes tactile sensations. One could therefore hypothesize that the ability of an organism to optimize the precision of its predictive errors may be crucial for both the robustness and malleability of its psychological self in relation to the world and particularly other people.

While we cannot do justice to these proposals and models and their many applications in this article, in the subsequent sections, we expand these insights into the mentalization of the body and the probabilistic, inferential model of the self to include fundamental developmental considerations regarding (1) the role of proximal interaction in the mentalization of the body and the minimal self; (2) the role of interoception and particularly affective touch in the mentalization of the body and the minimal self and lastly, (3) the role of action and ‘active inference’ (an important aspect of the Free Energy principle; see section 7) in the mentalization of the body and the minimal self.

4. The Touched Self: Multisensory Integration of Other Bodies and Embodied Mentalization

In this section we describe how interactions with other people are motivated and constrained by the same principles that govern the 'mentalization' of sensorimotor signals in the individual, and hence the mentalization of one's body can include signals from other bodies in physical proximity and interaction. As mentioned in the introduction, questions regarding minimal selfhood and the self-other distinction have received increasing empirical attention in developmental psychology over the past few decades. Some researchers have focused on multisensory integration and 'contingency detection' paradigms (Gergely & Watson, 1999 for review). The idea in these paradigms is to examine whether infants can distinguish between sensory changes arising from, and hence congruently with, their own motor actions (self) and sensory changes arising from the environment independent of their own action (non-self). For instance, some studies have now illustrated that infants as young as 3-5 months show sensitivity to body-related, proprioceptive-visual synchrony and, as motor control develops, also spatial congruency (Rochat & Morgan, 1995). In such paradigms, infants tend to respond differentially to visual feedback of their body parts moving synchronously and in spatial congruency to their own movements, rather than manipulated visual feedback that is asynchronous or spatially incongruent. These studies thus suggest that infants behave differently towards sensations that originate from their own body versus those that are caused by the environment.

Importantly, in these paradigms the critical variables that allows for this rudimentary self-other distinction are not modality specific. For example, the visual information of the body in the above infant experiments is the same between conditions. What differs is the 'amodal' property of temporal synchrony or spatial congruency between proprioceptive information (where infants felt the body to be) and visual information (where the infants saw the body to be). In simple terms, in the above experiment, it seems that infants can tell that what they see (via the visual modality) in front of them is similar or different from their own movements (detected via the proprioceptive modality). It seems therefore that such amodal ‘contingency-detection’ abilities are existent from early on and hence infants may be able to track the co-occurrence of sensory events across modalities. They may thus be able to progressively build predictions about the kind of sensorimotor signals that are most likely to occur together in relation to their embodiment. As we described in relation to the free energy model above, this may be the main building block of the self. Hence, the infants’ ability can be seen as evidence for an early, rudimentary self-other distinction.

Indeed, this is in fact the basis of most multisensory integration paradigms in adults: sensitivity to synchrony (the so called ‘glue of the senses’) across sensory modalities allows perceiving subjects to experience unitary multimodal events and to ascribe their origin on one’s own body or the external world. Accordingly, developmental studies on sensitivity to cross-modal synchrony have been considered as evidence for the early ability for a rudimentary distinction between self and other. In the preceding section, we described this progressive organization of perceptual input into distinct, unitary multimodal schemata as a process of ‘embodied mentalization’ and placed it in the context of a computational model of brain functioning (see Section 3 above). We are, in other terms, arguing that infants’ early ability to bind together sensory information in time and space lies at the core of a process of progressive mentalization of their embodied experience and hence at the core of the minimal self (even if rudimentary at this stage).

To this point, we have referred to the mentalization of the body as a process that concerns only the singular individual and its body. However, this analysis misses an important dimension of the mentalization process as it unfolds in development; namely, in early development caregivers’ bodies provide sensory data that can be plausibly experienced by infants as their own. Put crudely, the bodies of human caregivers provide an almost continuous embodied engagement in infancy, during which rich patterns of synchronicity and other forms of sensory and spatial contingency and congruency are frequently enacted. These seem to be reinforced by a rich repertoire of biological and cultural practices of interaction (e.g. hugging, kissing, singing, clapping, stroking, rocking, holding), as well as necessary and frequent routines of embodied engagements required to satisfy infant's basic biological and psychological needs (e.g. breastfeeding, washing, rubbing-cleaning, skin-to-skin sleeping, body-to-body temperature regulation and skin hydration, toileting; see also following sections). Extensive research has recently focused on 'visual' signals from other bodies, under 'mirroring' assumptions and theories of different kinds (see Gallese, 2013 for review). However, we propose here that the early, crucial role of such embodied practices in the formation of the minimal self is most obvious when one considers the special case of interpersonal 'touch' and the infant's immature motor system (see also next sections).

Indeed, there is even some evidence that fundamental feelings of body ownership develop based on early visuo-tactile integration mechanisms, before later processes of visuo-proprioceptive integration (Cowie et al., 2013). For example, a recent experimental study found that newborns look preferentially at visual face stimuli being touched in synchrony with their own face and were able to discriminate visuo-tactile synchrony from visual-tactile asynchrony (Filippetti et al., 2013). In this instance, the tactile stimuli in question were not caused by the infant (as her own movements on the above experiments). Instead, they were caused by another individual 'matching' the visual feedback for the purposes of the experiment. We propose that during proximal, caregiving interactions in early infancy, caregivers offer naturalistic 'matching' between their bodies and those of their infants. For example, a mother and a father may giggle playfully while tickling their baby. Tickling and giggling may thus be bound together as a frequent experience of the infant's body. More generally, we are proposing that what determines the early mentalization of one's own body, as opposed to that of another individual (i.e. which sensory input will be bound together and attributed to the self as most likely source of such input), may be somewhat paradoxically caused by embodied, social interactions. In other terms, embodied interactions between an infant and a caregiver may promote both self-other fusion (e.g. during moments of absolute synchrony, congruency or affective attunement) and self-other distinction (during inevitable moments of asynchrony, incongruency and non-attunement).

In everyday life, feeding, sleeping, calming down or entertaining routines typically include endless repetitions of multisensory bundles from at least two bodies (e.g. active and passive touch, proprioceptive and vestibular information, smell, temperature, visual and auditory feedback). During such experiences, the infant is therefore learning and responding to regularities and irregularities between the various sensorimotor 'bundles' of his and the caregivers body. As the baby learns to bind together 'own body' sensations from proprioception and vision regarding the position of her own limbs for example, she may also bind together sensations from proprioception, interpersonal touch and the smell of her father. Thus, the mentalization of the body, i.e. the progressive build-up of multisensory predictions in perceptual inference, *includes the body of the primary caregivers*. In this sense, the very first-person experience of my body as mine, as well as the building block of the self-other distinction, are constituted upon the presence of other bodies in proximity and interaction.

Before we go on, however, to further explain why physical contact and interpersonal touch has a unique, primary role in the minimal self, it is necessary to stress that, contrary to social constructivism (e.g. the idea that infants develop a mind because their caregivers have their mind in mind; see introduction), the critical social variable emphasized here is 'other *interacting bodies'*, rather than *other minds with higher order mental states*. We support of course the idea that the presence and behavior of caregivers is determined by their feelings and intentions towards infants, and these mental states (and the caregivers' psychological traits more generally) are of great importance for the degree of availability and exact behavior of caregivers (see also Section 8 below). However, the constitutive factors of minimal selfhood we are trying to account for in this section concern the mere embodied presence of caregivers and some minimal caregiving activity towards infants. For example, while the presence and behavior of a mentally absent, versus a highly responsive caregiver would be different and may have different effects on the infants' personality and of course their mentalizing, we argue here that infants would develop a minimal self of *some* quality regardless of such differences.

Let us illustrate the above point in reference to a common experience in parenting, paired with recent empirical study on infant holding. Most parents would recognize the fact that it is far easier to calm a crying baby while standing up and pacing in the room with the baby in one's arms, rather than by holding the baby in one's arms while seated. In agreement with this, a recent study in infants less than six months old found that being held and carried by a walking mother led the infants to immediately stop voluntary movement and crying and exhibit a rapid heart rate decrease, compared with holding by a sitting mother (Esposito et al., 2013). Furthermore, similar motor, vocal and heartrate 'calming' responses were observed in mouse pups, supporting the idea of a conserved embodied component of mammalian mother-infant interaction. We assume that the mental states and 'mindreading' capacities of mothers towards their babies do not typically vary depending on whether they are walking or sitting, and yet it appears that the particular bodily behavior of the caregiver (pacing versus sitting) seems to have direct behavioral and physiological effects on infants. We thus suggest that these findings demonstrate the importance of embodied caregiver-infant interactions *per se* (e.g. mobile versus static holding), without the need to refer to the sharing of any higher order, mental, or even spatial, concepts such as intentionality, empathy or perspective. In short, embodied interactions with caregivers may have a fundamental role in shaping the mentalization of one's own body, even prior to the more complex psychological considerations and exchanges between infant and caregiver such as those described by classic mentalization and theory of mind accounts (see introduction).

Similar effects of embodied and primarily tactile interactions between parents and their offspring have long been established in other mammals (Harlow & Zimmerman, 1958; Panksepp, 1998). For example, in primates, pro-social tactile stimulation (i.e. licking and grooming) attenuates responses to stress, with beneficial long-term effects (Korosi and Baram, 2010; Weaver et al., 2004) and activates endogenous analgesic processes mediated by opioid (Kehoe & Blass, 1986) and oxytocinergic mechanisms (Agren et al., 1995). The involvement of these neurobiological pathways, implicated in stress and pain regulation (e.g., Lundeberg et al., 1994), as well as the formation and maintenance of close social bonds (Insel, 2000; Nelson and Panksepp, 1998) provides some indirect support for the idea that social touch is critical for the development of the affective nucleus of the self. This idea is also supported by clinical and developmental studies suggesting that touch-based interventions such as massage and 'skin-to-skin' contact can have positive psychological and physical effects in preterm infants and children (Feldman and Eidelman, 2003; Field et al., 2010 for a review) and in adult illness (e.g. Hart et al., 2001). Unfortunately, such studies have methodological limitations and relevant systematic research in human infants is sparse and mostly correlational (e.g. Sharp et al., 2012).

Moreover, for the most part, scholars of human infancy, including some psychoanalysts, tend to claim that such effects in humans are mediated by parents' mental states and related, higher-order psychological concepts (e.g. theory of mind, attachment style, etc.). Even in theories that have stressed embodied aspects of infant-caregiver relationship, e.g. 'affect attunements' (Stern, 1985), "contingency detection module," (Gergely & Watson,

1999) or ‘marked mirroring’ (Fonagy et al., 2002), these are rather quickly embedded in more complex mentalistic conceptualisations about the caregiver’s mind. For instance, there is the assumption that infants’ minds are first ‘read out’ by the mothers and then responded to accordingly (Fonagy et al., 2002), or alternatively, the infant’s ability to distinguish itself from others, precedes the influence of such attunements on the self (Stern, 1985).

While we do not deny the role of such forms of relatedness and intersubjectivity, we agree with philosophers such as Zahavi (2015a) and neuropsychoanalysts like Mark Solms (2013) that such factors are secondary to the mechanisms responsible for the minimal self (see also Ciaunica & Fotopoulou, 2016). Nevertheless, we propose that *proximal embodied interactions* of caregiving (as opposed to more higher-order, mentalistic exchanges) contribute to the constitution of the self from the onset, so that the primary feelings of ‘selfhood’, or ‘mineness,’ that accompany all subjectivity are deeply rooted in social interactions. We further specify below why such embodied interactions are especially relevant and necessary for the formation of minimal selfhood, particularly as regards how interoception is mentalized by mechanisms of multisensory and sensorimotor integration (perceptual and active inference in the terminology of the free energy framework).

5. The Felt Self: The Mentalization of Interoceptive Signals

What kind of bodily signals become ‘mentalized’ to form the basis of minimal selfhood? Although it has long been proposed that bodily self-awareness relies on an integrated representation of multiple streams of sensory and motor information as described above, there has been a strong bias in the kind of bodily signals studied. Specifically, most scientific investigations have focused on ‘multisensory integration’ paradigms that study the integration of exteroceptive (e.g. vision, audition, touch) signals, or on sensorimotor integration paradigms that may also include motor, efferent signals and proprioceptive or vestibular feedback. Remarkably, however, until recently little work on bodily self-awareness concerned interoception.

According to a recent reclassification of the senses, interoception refers to the perception of the physiological condition of the body, involving modalities such as temperature, itch, pain, cardiac signals, respiration, hunger, thirst, pleasure from sensual touch, and other bodily feelings relating to homeostasis (Craig, 2010; Critchley et al, 2004). It is distinct from the exteroceptive system, which refers to the classical sensory modalities for perceiving the external environment (e.g. vision, audition), as well as proprioceptive, vestibular and kinesthetic input informing about the movement and location of the body in space (Blanke & Metzinger, 2009; Craig, 2010; Critchley et al, 2004). Crucially, contrary to classic views of interoception as 'the perception of the body from within', the current notion of interoception is tightly linked to homeostasis; interoceptive signals are considered crucial in informing the organism regarding the homeostatic state of the body in relationship to experiences originating from within the organism (e.g. cardiac and respiratory functions, digestion, hunger, thirst), or outside it (e.g. taste, smell, affective touch, pain). One might say that interoception informs the body about *how* the body itself is doing in relation to certain inherited, homeostatic needs (e.g. one may be dehydrated, or stung by an insect), while exteroception informs the body about environmental changes in relation to such needs (there is a river ahead) but independently of the physiological state of the body itself. Moreover, interoception is thought to rely on separate specialized neuroanatomical systems that are associated with the autonomic nervous system, special spinal cord pathways and subcortical and cortical brain areas mapping homeostatic and motivational states (Critchley et al., 2004; Damasio, 1994; Panksepp, 1998). Moreover, recent work has suggested that interoception is uniquely related to the generation of subjective feelings, informing the organism regarding its levels of arousal and bodily needs (Craig, 2009; Seth, 2013). As such, the impact of interoception is thought to extend beyond homeostatic (and allostatic) regulation.

Interoception has been ascribed a central role at the heart of self-awareness (Craig, 2009; Critchley et al, 2004; Damasio 1994). In relation to the free energy framework, it has been recently proposed that subjective feeling states arise from predictive inferences on the causes of interoceptive signals (Seth et al, 2012; Seth, 2013; Barrett & Simmons, 2015; Pezzulo et al., 2015). These ‘‘interoceptive predictive coding’’ models are compatible with the so-called James-Lange theory of emotions which links feelings with the perceptions of physiological changes. Classic debates in psychology have unfolded about whether bottom-up, direct bodily signals and/or top-down cognitive evaluations of physiological changes are responsible for feeling states. The advantage of interoceptive predictive coding models is that they can specify the dynamic balance between bottom-up interoceptive predictions errors and top-down cognitive signals, exactly as explained with regards to exteroception and proprioception. The disadvantage of interoceptive theories is that they may be missing other, more fundamental aspects of emotional consciousness that may relate to the optimization of precision in both perception and action (see Fotopoulou, 2013), or the optimization of the free energy itself (Joffilys & Coricelli, 2013). Unfortunately, this aspect of the model escapes the scope of the current article.

Importantly for the present purposes, there is also associated, preliminary evidence that interoception can uniquely shape the bodily self, as studied in multisensory integration paradigms and neuropsychiatric disorders (see Barrett & Simmons, 2015; Seth, 2013; Tsakiris, 2016 for reviews). For example, participants with lower abilities to detect their own heartbeat seem more susceptible to bodily illusions of synchronous visuo-tactile stimulation (Tsakiris, Tajadura-Jimenez & Costantini, 2011; Tajadura-Jimenez & Tsakiris, 2014).

How can one reconcile this view with more classic considerations of the constitution of the minimal self and related debates on intersubjectivity as outlined above? At first sight, the potential role of interoception in the minimal self may be interpreted as evidence in favor of the idea that the sense of self primarily arises from the individual's experience of his or her own body. The resulting inner feelings of 'arousal', 'wakefulness', 'wellness', or lack thereof, combined with exteroceptive and motor signals regarding the body, could thus be fully sufficient to form the basis of subjectivity and the self, and a fundamental source of information regarding the self-other distinction. This is indeed the view that several scientists have put forward recently (Craig, 2009; Damasio, 1994; Critchley et al., 2004; Seth et al., 2012).

Upon closer inspection however, it appears that interpersonal interactions are necessary in shaping the mentalization of interoception, and not the other way around. This claim is supported by two main observations, one in relation to perceptual inference and the other in relation to active inference. We will unpack these observations in the following two sections, respectively. Firstly, interoception itself derives from the outside and other bodies as much as from the inside of the body. Secondly and perhaps most importantly, in early infancy, when the motor system is not yet developed, interoceptive function and homeostasis are wholly dependent on embodied interactions with other bodies.

6. The Affectively Touched Self: Learning Bodily Pleasure and Pain in Social Perceptual Inference

As aforementioned, contrary to classic views on interoception, contemporary accounts define it as the set of modalities informing the organism regarding the homeostatic state of the body, both in relationship to experiences originating from *within* the organism (e.g. cardiac awareness, hunger), *or outside* it (e.g. taste, smell, affective touch, pain). In this article, we will focus on affective touch, as a key example of how proximal, embodied interactions affect the minimal self.

Recent neurophysiological, neuroimaging and behavioural studies suggest that certain tactile experiences, such as gentle caress-like strokes, are processed by at least two separate neurocognitive systems. First, as it has been known for decades, tactile stimuli are processed in terms of their exteroceptive, discriminatory processes in classical peripheral pathways and somatosensory cortical areas. Second, it was recently demonstrated that a specialized peripheral and central system codes for the *affective* properties of the same tactile stimulus (Löken, Wessberg, Morrison, McGlone, & Olausson, 2009). Whereas purely sensory touch is conveyed by skin mechanoreceptors projecting to the thalamus and primary somatosensory cortex, the neurophysiological system for affective touch (Vallbo, Olausson, & Wessberg, 1999) seem to rely on a distinct subgroup of mechanoreceptors: tactile C-fibres (CT), responding only to slow (between 1-10 cm/s), caress-like touch and leading to subjective pleasantness (Löken et al., 2009). Crucially, CT afferents take a distinct ascending pathway from the periphery to a different part of the thalamus and then to the posterior insular cortex (Morrison et al., 2011). According to some researchers, the latter pathway mediates an early convergence of sensory and affective signals about the body, which are then re-represented in the mid- and anterior insula, the proposed sites of interoceptive awareness (Craig, 2009; Critchley et al., 2004).

Thus, while gentle, stroking-like touch originates from outside the body, it appears to simultaneously convey information about the ‘inner’ body (e.g. reductions in physiological arousal) and the external world (e.g., ‘sensation of skin stimulation of some density and softness characteristics, slow speed and little friction). Moreover, the CT-system has been found to respond optimally to touch of human temperature rather than colder robot-based touch (Ackerley et al., 2014). This and related findings (for review and the so-called ‘social touch' hypothesis see Morrison, Loken, & Olausson, 2010) suggests that this system may be specialized not only for processing affective touch, but also specifically social affective touch. Crucially, a recent study found that nine month old infants are sensitive to the particular physical properties of affective touch: CT-optimal but not non-optimal velocities of tactile stimulation led to heart rate decelerations in the infants, possibly reflecting relaxation and increases in their behavioral engagement (gaze shifts and duration of looks) with the stroking stimulus (Fairhurst et al., 2014). Thus, this type of affective, social touch may be another example of a specific, embodied social behavior that can regulate homeostasis and influence the basic, feeling states of the infant.

Moreover, we need to highlight here that, as aforementioned, any instance of such gentle, slow touch (a slow, gentle caress by a mother) simultaneously activates pathways (the CT system) relating to interoception and the physiological state of the body (e.g. reductions of physiological arousal), as well as some features of the external stimulus touching the infant (e.g. via fast conducting, myelinated fibers). We thus speculate that given that these interoceptive and exteroceptive data generated by a single tactile stimulus occur by necessity at the same time and place (thus they are characterized by the amodal properties of temporal and spatial congruency that bind modalities together and create predictions and generative models of their causes), they are progressively mentalized as one experience. In other terms, the caregiver's touch contributes directly to the integration of the infant's interoceptive and exteroceptive bodily experiences, possibly helping the infant experience the skin as the boundary between her body and external world. Thus, social touch, an essential part of early mother-infant interactions, may have a unique developmental role in progressively establishing the physical boundaries of the psychological self.

Unfortunately, to our knowledge, no systematic, developmental studies have focused specifically on the role of affective touch in the formation of the minimal self. However, the aforementioned application of multisensory integration paradigms to the study of infant body perception (e.g. Filippetti et al., 2013) suggests that the specific, developmental role of affective, social touch should be studied experimentally in early infancy and childhood. Moreover, indirect confirmation can be found in studies on neurodevelopment in low birth weight infants (see Gallace & Spence, 2010 for review). For example, Weiss and colleagues (2004) observed that infants of mothers who used more stimulating touch during feeding at 3 months had better visual-motor skills and more advanced gross motor development. Other evidence has shown that skin-to-skin contact following Caesarean section may help maintain temperature of newborns and reducing new-born stress (for review, Stevens, Schmied, Burns, & Dahlen, 2014). Studies with older infants have showed that the infant's arousal can be regulated by soothing, touch-based behaviors by their caregivers, in addition to self-soothing behaviors (e.g. thump sucking) (Beebe & Lachmann 1998; Rothbart, Ziaie, & O'Boyle, 1992). Taken together these findings suggest that there are a set of embodied interactions between infants and caregivers, such as gentle touch of the skin, that have direct effects on the infant's physiological arousal. In this sense, certain interpersonal interactions seem capable to 'bind' together inner feelings about the state of the body with external perceptions of the body and the world.

In addition, indirect confirmation of our suggestion comes from studies on adults. Affective touch has been shown to provide information about the emotions and thoughts of other individuals, i.e. the touch providers (Hertenstein, Keltner, App, Bulleit, & Jaskolka, 2006) and the touch receivers (Gentsch, Panagiotopoulou & Fotopoulou, 2015). More specifically as regards to the minimal self, a series of recent studies focused on the role of affective touch in the sense of body ownership. Using the RHI paradigm, three independent studies (the first of these being from AF's lab), have now found that slow, caress-like touch of CT-optimal velocities and properties enhanced various subjective and behavioural measures of the RHI more than fast, emotionally-neutral touch (Crucianelli et al., 2013; Lloyd et al, 2013; van Straleen et al., 2014). That is, affective, pleasant touch delivered by another individual appears to be a strong determinant in the process of multisensory integration that determines how a body part is subjectively experienced as mine. In keeping with this observation, several studies on patients with clinical disorders of the minimal self, such as body ownership disturbances, demonstrate that affective touch increases a sense of body owernshipownernshp, such as patients who at least momentarily accept their disowned arm as theirs following affective touch (see Gentsch et al., 2016 for review).

The effects of embodied interactions such as affective touch on the minimal self can be drawn in relation to pain, including cutaneous noxious stimulation generated from the outside. Indeed, affective touch and cutaneous pain are two sub-modalities of interoception with contrasting affective qualities (pleasant/unpleasant) and social meanings (care/harm). In experimental studies, it is well established that social support can modulate psychological and neurophysiological response to pain, in adults and in children (see Decety & Fotopoulou, 2015; Krahé et al., 2013, for reviews). Moreover, in experimental and neuroimaging studies with adults, we (AF's lab) have shown that this pain modulation depends on particular 'embodied' social support variables (e.g., the presence of another individual, affective touch by another individual), as well as individual differences in the perception of social relationships themselves, namely attachment styles (Hurter et al., 2014; Krahé et al., 2015; Sambo et al., 2013; Krahé et al., 2016). Insecure attachment styles in particular (characterised by negative expectations of social support), which may be linked with an impoverished oxytocin system (see Uvnäs-Moberg, Handlin, & Petersson, 2014), seem to moderate the relation between social support and pain (see also Meredith, 2013). Higher attachment anxiety (associated with seeking and craving signs of reassurance) led to reduced pain in the presence of a high vs. low empathic stranger (Sambo et al., 2013) and to reduced pain when receiving CT-optimal, affective touch (Krahé et al., 2016), while higher attachment avoidance (associated with distancing from others and preferring to cope alone) led to increased pain in the presence of a stranger (Sambo et al., 2013) or one's romantic partner (Krahé et al., 2015) and when receiving affective touch (Krahé et al., 2016).

Thus, in adulthood, it appears that embodied social interactions strongly influence our experience of bodily pain, subject to our predictive models about the availability and support of others (operationalised in the above studies as attachment models, presumably built during childhood). We have proposed elsewhere (Decety & Fotopoulou, 2015; Krahé et al., 2013) that from the perspective of the free energy model, this modulation may be conceptualised as a modulation of the precision (salience) of nociceptive signals in a social context. In other words, our perception of pain, and of bodily threat more generally, may vary not only according to how much tissue damage is communicated by nociceptive, peripheral pathways, but also according to how much social support we predict is available to us in a given situation, or more generally. If we are inclined by prior experiences to trust others and their potential active help during bodily threat, we may experience and react to pain-related prediction errors differently than when we are not trusting the availability or effectiveness of others' support. These results seem to provide an answer to the question we raised in the beginning of this article regarding the paradoxical observation that an experience of bodily harm in childhood invites immediate social attention and reaction. The experience of pain, so intimately connected with subjectivity and an individual's own body, may actually be shaped by interpersonal attention and other social factors.

Taken together, the studies reviewed in this section suggest that even the subjective feelings of pleasure and pain regarding one's own body, which we may conventionally think of as being purely arising from within ourselves, are actually formed in interaction with other bodies and minds in development and in adulthood. We accept, of course, that evolution has equipped humans with specialised systems for bodily pleasure and pain, as means to ensuring homeostasis (see also Solms, 2013). However, we are suggesting that in the developmental transition (both phylogenetic and ontogenetic) from physiological reaction (e.g. nociception) to subjective affect (e.g. pain), these modalities have been progressively 'mentalized' to form predictive, psychological feelings and action tendencies (e.g. pain-related expressions) that carry with them the stamp of proximal interpersonal interactions. As we discussed in Sections two and three, these very feelings are considered as the core of our minimal self. Thus, it appears that the core of selfhood is interpersonally constituted.

7. When the Motor System is Not Yet Developed: Active Interoceptive Inference via the Other's Body

The above assertions on the affective core of the minimal self require an additional consideration. In this section, we put forward the radical claim that because human infants are born without a fully matured motor system, and hence they cannot regulate their own homeostasis unaided, the actions of their caregivers necessarily determine how they come to experience the affective core of their embodied selfhood. As we described above, in the free energy framework, action is understood as the reverse process of perceptual inference: we selectively sample the sensory inputs that we expect (i.e. by fulfilling proprioceptive predictions) in order to add accuracy to our predictions about sensory states (Friston, 2010).

However, as in the case of perceptual inference, the conceptualization of active inference can be enriched by fundamental developmental considerations. Specifically, action and perception do not mature at the same time and this imbalance has particular implications for the mentalization of interoception. Given an infant's immature motor system, what kind of models can she form based on active inference and sensorimotor integration? Young infants have at their disposal a number of reflexes, including autonomic ones, and they show some movement of the head, including the face and eyes, the limbs and the trunk. Thus, for example, when infants open their eyes, or turn their heads, they can progressively learn to perceive changes in lighting. Similarly, reflective and purposeful hand or leg movements may frequently be met with some obstacle in the world, e.g. a blanket, a toy that makes sounds, etc., that is exteroceptively perceived. These first unaided sensorimotor experiences afford several opportunities to the infant to progressively build generative models regarding the possible causes of their sensory states in the external world. For instance, an infant can learn that closing her eyes or looking away causes changes in her visual input, which she can then learn to implement when large changes in environmental light occur. However, infants lack strength and control in their large antigravity muscles and are helpless in supporting their own weight, and are of course unable to initiate and execute, complex sequences of purposeful movements. Therefore, a young infant cannot position and balance itself, feed itself, thermoregulate, or protect itself from tissue damage (e.g. skin burns, bone fractures etc.).

Thus, in the case of such interoceptive modalities, *no movement on the part of the infant alone, can change certain key neurophysiological states relating to homeostasis*. In terms of the free energy principle (see Pezzulo et al., 2015 for a detailed consideration of the relation between active inference, homeostatic and allostatic regulation), as far as these interoceptive modalities go, there can only be minimal implementation of interoceptive predictions by simple autonomic or motor reflexes (e.g. sucking the breast, or the bottle, sweating, or shivering). Put in simple terms, the infant can suck a dummy, or their fingers to regulate some of their neurophysiological states, but in order to eventually change states related to food intake and hence eventually build generative models regarding hunger and satiation, *someone else* needs to be in proximity and interaction and to offer the breast or the bottle. Similar considerations apply for noxious stimulation, thermoregulation and other domains of interoception. Thus, the unaided infant cannot use action to collect evidence about the causes of its interoceptive experiences and thus it cannot test its interoceptive predictions against the world.

Instead, the infant's autonomic and motor reflexes in response to unpredicted physiological states (e.g. crying and kicking when hypothalamic function detect that glucose level are not within the predicted viable range) can elicit the attention of the caregiver and ensure that the caregiver tries to change the physiological state of the infant (for example by feeding it) until the homeostatic needs are met (i.e. glucose levels are within the predicted range). Thus, updating interoceptive predictions in infants includes information regarding the reaction of caregivers to infants' initial autonomic and proprioceptive predictions. Therefore, it is the adult's actions and reactions, their frequency and multisensory characteristics that will generate changes in interoceptive states and hence ultimately contribute to the 'mentalization' of physiological states in the infant. Accordingly, in this section, we put forward the radical claim that it is exactly because a human infant depends on the caregiver to regulate her homeostasis that the interaction with the other is woven into the very emergence of the self. In the terminology of the Free Energy model, the origins of interoceptive active inference are always, by necessity social, and thus core subjective feelings such as hunger and satiation, pain and relief, cold or warmth have actually social origins.

Moreover, in good-enough caregiving environments, such caregiving behaviors are met, not only by facial expressions and other "mentally" attuned responses, but – and crucially – with a variety of proximal, *embodied* responses, such as soothing touch, holding, feeding, dressing etc., which, as we described above, can themselves produce further changes in the infant's physiology (e.g. heart-rate reductions). Basic caregiving behaviors therefore do not only 're-sample' the world on behalf of the infant's interoceptive predictions, they are at the same time the source of multisensory input (e.g. auditory, tactile, olfactory, and visual bundles of experience) and amodal properties such as rhythm, frequency, synchrony and other such variables (see also Stern's vitality affects, 1985). Other bodies and their actions therefore stand right in the midst of all embodied mentalization processes (the formation of generative models based on perceptual and active inference) of the infant.

It follows that the progressive mentalization of the affective core of selfhood does not take place by processes that belong to the singular infant, as certain theories assume (e.g. Damasio 2010; Craig 2009; Solms, 2013; Seth et al., 2012; Seth & Friston, 2016). Instead, active and perceptual interoceptive inference in development is by necessity mediated by the actions of caregivers that bring about physiological changes, and hence shape the perception of bodily satisfaction, relief, pleasure, pain, or lack thereof.

Freud had wrongly assumed that the human infant's survival needs is the ultimate motivation for our early social-relating. Seminal studies have since established that humans have an innate social attachment drive, unrelated to hunger or thermoregulation, and a corresponding lifelong need for social connection (Bowlby 1969; Harlow & Zimmerman 1958; Panksepp 1998). Contrary however to the recent emphasis on mentalistic concepts such as 'attachment styles,' it is useful to remember that earlier proponents of this view indeed emphasized the embodied rather than the mentalistic dimensions of this drive, such as the 'need for physical proximity' (Bowlby, 1969) and 'contact comfort' (Harlow & Zimmerman, 1958). Indeed, we believe that the primacy of our social attachment drive should not obscure the important, embodied role of caregivers in regulating the infant's interoceptive states and in turn, the foundations of the minimal self. The drive to stay close to caregivers may be an important inherited prior. Separation from caregivers therefore may elicit physiological reactions to infants that can be satisfied only by embodied, proximal interactions like the ones we described in this and previous sections. Thus, we are suggesting that *the origins* of all mentalization processes are not only embodied but also by necessity involve other people's bodies, their physical presence, proximity, contact and most importantly, their homeostatically-relevant action.

Of course, the above conclusions regarding the embodied and social origins of our self-awareness do not mean to deny the importance of later, more higher-order and 'mindful' interactions between children and their caregivers. We just propose that these are psychological and social extensions of more fundamental, biologically prescribed processes of embodied care in early infancy. Thus, we briefly trace the progressive development of the self and the self-other distinction in the final section below.

8. The Maturation, Individuation, and Strengthening of the Self: Interoceptive Awareness and Social Cognition

The Development of Interoceptive Awareness. In development, as ongoing intersubjective bodily interactions with the caregiver get more complex, children build increasingly more sophisticated models of their own interoceptive states, as well as strategies for minimizing free energy in the interoceptive systems. This constitutes a capacity for awareness of interoceptive states, which seems to play a fundamental role in mental health. Indeed, the concept of interoceptive awareness has recently become central to interdisciplinary research on self-awareness.

Interoceptive awareness reflects our ability to become aware of internal states, and is typically operationalized in psychophysiological research by measures of "interoceptive accuracy" (IAcc), such as, for example, one's ability to accurately feel and count heartbeats during short intervals (Schandry, 1981),or to detect the degree of synchronicity between individual heartbeats and auditory tones (Brener, Liu & Rign, 1993). Similar methods have been applied in other interoceptive systems, such as respiratory or gastric sensitivity. IAcc is considered a trait-like proxy for interoceptive awareness (Garfinkel, Seth, Barrett, Suzuki, &

Critchley, 2015); inter-individual differences in performance allow us to distinguish between people with higher vs. lower levels of IAcc. Individual differences in cardiac IAcc have been linked to mental health (Herbert & Pollatos, 2012), with very high IAcc predisposing to anxiety, while symptom severity in patients with alexithymia is inversely related to IAcc. Low IAcc is also associated with depersonalisation, personality disorders, psychosomatic complaints, and eating disorders (Pollatos et al., 2008). In healthy adults, research into IAcc has been almost exclusively concerned with emotion, and is associated with the intensity of emotional experience and emotion regulation (Critchley & Nagai, 2012). For example, individuals with high IAcc are more able to self-regulate (Herbert & Pollatos, 2012) and tend to follow their intuition more in decision-making tasks (Dunn et al., 2010). In sum, the available evidence suggests that interoceptive awareness is important for emotional awareness and mental well-being.

Interoceptive functions supporting homeostasis, such as vagal tone, have been extensively investigated in young children. For example, as with adults, vagal tone (VT) is linked to children's ability to regulate arousal and similarly underlies individual differences in emotion-regulation and temperament. The first months post-partum are characterized by relative instability of key cardiovascular variables such as heart rate variability (HRV) and VT, which become moderately stable by the end of the first year (Fox, Schmidt, Henderson, & Marshall, 2007), due to physiological maturation (Fracasso, Porges, Lamb, & Rosenberg, 1994). Importantly, their *levels* depend on the quality of caregiving they receive (McLaughlin et al., 2015), such as parent-infant contingency during interaction (Feldman, Magori-Cohen, Galili, Singer, & Louzoun, 2011; Feldman, 2007), and are predictive of self-regulation abilities in three-year-old children (Fracasso et al., 1994). Attachment style has been shown to shape infants' response to stress and environmental challenges and, as with adults, studies have replicated the importance of high VT for children's emotional regulation and well-being (Hill-Soderlund et al, 2008).

As we argued earlier, during embodied interactions, the multisensory input that the carer provides to the infant not only facilitates the perceptual inferences needed for the infant's body-ownership, but also provides the means by which the infant can learn to identify/become aware of her distinct internal needs. There is no reason to assume that the infant has a clear awareness of differentiated interoceptive states from the outset – neither does the parent. Moreover, given the infant's limited behavioural repertoire, it rests upon the carer to firstly respond as consistently as possible with her instinctual behaviour to the infant's embodied expression (e.g. crying) and eventually learn to detect the precise interoceptive need that her multisensory embodied interactions with the infant should aim to settle (e.g. learn at which times crying is more likely to be associated with hunger rather than thermoregulation needs). Contingent, appropriate, "good enough" (Winnicot, 1972) responses by caregiver to the interoceptive needs will enable the infant as well as the carer to generate more accurate models of interoceptive inferences that, we propose, will eventually lead to higher interoceptive awareness. However, given the lack of empirical data on the development of interoceptive awareness during the first few years of life, our hypothesis remain to be directly tested in longitudinal developmental studies.

From synchrony to separation. In this target article we have also argued that interoception and in particular the mentalization of interoceptive signals play a critical role in self-other boundaries. The distinction between self and other, which is crucial for self-awareness, *is equally essential* for awareness of other people, as the brain must monitor whether sensations, events, and mental states should be attributed to oneself or not. Correctly identifying the origin of bodily and mental states is necessary for social relatedness. For example, how can I share the pain of another individual without forgetting it is not my pain?

The process by which models that enable us to mentalise our body are built up have clear implications for our understanding of the dynamics of self-other boundaries, and of their dependence on multisensory, sensorimotor and interoceptive signals.

Multisensory signals have been extensively studied in the context of body-ownership as we described earlier. Recent experimental studies have extended the basic mechanism of multisensory synchronous input that was typically used to elicit changes in body-ownership to study self-other boundaries. For example, in the enfacement illusion (see Tsakiris, 2016 for a review), watching another person's face being touched synchronously with one's own face evokes changes in self-face recognition, so that we perceive the other person's face as more similar to our own: participants perceived the other's face as physically more similar to their own (Sforza et al., 2010; Tajadura-Jimenez, Grehl & Tsakiris, 2012; Tsakiris, 2008); rated the other as conceptually closer to themselves, and also ascribed more self-like personality traits to the other (Paladino, Mazzurega, Pavani and Schubert, 2010); and displayed increased emotion recognition of the other's emotional facial expressions (Maister, Tsiakkas & Tsakiris, 2013). These findings show how multisensory signals may blur self-other boundaries.

Therefore, from a predictive coding point of view, perceptual inferences of the kind described in the multisensory literature and bodily illusions (and in the absence of agentic motor actions) may provide the basis for bodily self-awareness (see RHI), but at the same time may also create the conditions for blurring the boundaries between the self and other (see the enfacement illusion). Paradoxically, therefore, the early carer-infant interactions that rely largely on multisensory synchrony and contingency provide the basis upon which the infant starts building a representation of its own body that is largely shared with the other. Psychoanalytic insights are particularly relevant here for understanding the transition from the blurring between self and other towards a self-other distinction essential for the consolidation of second person perspective. Indeed, responsiveness and mirroring are never perfect ; they are ‘good enough (Winnicot, 1972) and ‘marked’ (Fonagy et al., 2002). Marking, frustrations, incongruencies and inhibitions in intersubjective interactions progressively strengthen the self-other distinction, as psychoanalysis has long pioneered.

Importantly, as we have argued throughout this paper, as one’s body is not simply perceived from the outside (i.e. exteroceptively, as when we look at the mirror) - it is also felt from the inside (i.e. interoceptively, as when one feels her racing heart) – interoceptive signals and their awareness may play an important role in self-other boundaries. In this regard, Tsakiris et al. (2011) observed a negative correlation between levels of interoceptive awareness and the changes in the experience of body-ownership (see RHI) and self-identification (see Enfacement Illusion). People with lower levels of interoceptive awareness showed a stronger change in body-ownership following exteroeceptive stimulation, suggesting that, in the absence of accurate interoceptive representations, one’s model of self is predominantly exteroceptive (see also Schauder et al., 2015). The less sensitive one is to internal states, the more influence external stimulation appears to have. Interestingly, lower levels of interoceptive awareness were correlated with greater self-other blurring during the enfacement illusion (Tajadura-Jimenez et al., 2012; Tajadura-Jimenez & Tsakiris, 2014). In the context of the predictive models of self-processing described earlier, such findings highlight that *perceptual* inferences resulting from multisensory perception update self-representations by blurring self-other boundaries, especially when interoceptive awareness is low.

What would then be the role of greater interoceptive awareness in facilitating clear self-other boundaries? Within a predictive coding account, the active, rather than purely perceptual, inferences may play a key role as they would more accurately predict sensory states. Past literature from experimental psychology has aptly highlighted the critical role that motor voluntary actions and our sense of agency over them play for self-other distinction. The sense of agency is a fundamental dimension of embodied experience that describes not simply the experience of *having* a body but also of *controlling* one's body in order to cause desired effects in the environment (Haggard & Tsakiris, 2009). There are two reasons why interoceptive awareness may play a vital role in the sense of agency. First and foremost, biological organisms depend on homeostasis, and many voluntary actions have the specific aim of ensuring homeostasis (e.g. looking for food when hungry). As such, the action system and its experiential dimension (i.e. agency) should have clear input links from the interoceptive system. Second, our actions produce exteroceptive effects in the world but they also have interoceptive consequences. Ainley, Sel and Tsakiris (under review) tested this hypothesis by examining the relation between levels of interoceptive awareness and a well-established measure of agency, namely the "intentional binding" effect. In intentional binding, the perceived time-awareness of voluntary actions and their effects are perceptually attracted, so that actions and effects are perceived to occur closer together in time than they actually do (Haggard, Clark, & Kalogeras, 2002). Ainley and Tsakiris showed that participants with higher levels of interoceptive awareness showed a larger intentional binding (that reflects a stronger sense of agency). Therefore, our experience of agency, critical for self-other distinction, may also depend on the role that interoception plays in delineating the self (Craig, 2010; Damasio, 2010; Seth, 2013), in line with the view that active inference increases the accuracy of our predictions about sensory states that may be the result of our own or other's actions and bodily states.

Interoceptive Awareness and Social cognition. Lastly, the correct identification of the origin of bodily and mental states is necessary not just for self-other distinction but also for social relatedness and cognition. Emotional contagion, mimicry, body resonance, perspective-taking, and theory of mind have become key topics in the prolific field of social cognitive neuroscience and have been used to operationalize different facets of empathy, which is one of the hallmarks of social relatedness (Bernhardt & Singer, 2012; Decety, 2011). An important unresolved issue is the question of "self-other" overlap (Preston & Hofelich, 2012). Simply put, "self-other" overlap is thought to arise when an observer engages in an isomorphic state (e.g. the same emotion) to the person observed. To what degree can we distinguish between self and other at the very time that we are trying to understand each other?

Interoception may have a unique dual role in this respect. As we outlined in previous sections, its origins are social. Yet, as children develop interoceptive awareness of their own body as an object of their perception through interactions with the other, interoceptive awareness also acts as a constituting element of the self that safeguards against excessive self-other blurring. This view paves the way for a radically new approach to the question of self-other relatedness. According to recent models of social cognition, the default modus operandi of the social brain is to represent one's own self (e.g. one's own perspective, emotion, beliefs etc.). Switching from self to other to achieve a partial co-representation of self and other is therefore an effortful process that, at least to some extent, requires the attenuation of self-representations (Bird & Vidding, 2014). In psychoanalytic terms, one needs to inhibit the original self-other blurring that constituted the self in order to be able to focus on the other as an independent, separate object of perception. From the interoceptive point of view, the hypothesized attenuation of self-representations so that the other is better represented would need to be extended to interoceptive feelings. According to this view, lower levels of interoceptive awareness may provide an advantage in switching from self to other because the attenuation of interoceptive prediction errors may be computationally easier. Consider the case of emotion contagion, where exposure to someone else's emotion brings about a similar affective state in the perceiver but without explicit awareness that the catalyst for the state should be attributed to the other individual. A lack of awareness of the origin of the affective state may indicate low IAcc. An alternative hypothesis that we put forward holds that understanding others requires a ‘good enough’ representation of one’s own (interoceptive) states, because our representation of the other’s states is based on an awareness of how *their states affect us (on the basis of how they affected us originally)*.

Moreover, this self-representation should now display sufficient stability to prevent the blurring of self and other. Such blurring may indeed happen in adulthood, particularly in certain moments of emotional turmoil, which may arise in attempt to maintain either bodily integrity or attachment, and thus may be more common in some psychopathologies such as borderline personality disorder (Fonagy et al., 2002), or neuropathologies such as somatoparaphrenia (Fotopoulou et al., 2011). In the case of empathy, which, unlike emotion contagion, requires explicit knowledge of the origin of the emotion (Bird & Vidding, 2015), it is as yet empirically untested *whether one needs an accurate sense of one’s own body, in order to emphasize with the other, as a separate individual*.

9. Conclusion

Building mainly on empirical research on affective touch, pain and interoceptive awareness, we have argued in favour of a reconceptualization of minimal selfhood that stresses that the fundamental principles of the human *embodied* condition determine both the embodied and the relational origins of the self. Specifically, we have described as ‘embodied mentalization’ the process of building generative models by detecting regularities and irregularities in modality-specific and “amodal” properties, such as synchrony, and organizing sensory input of both personal and interpersonal origins into distinct, unitary multimodal schemata (perceptual inference). We have also stressed that such models refer not only to exteroception but also to interoception, the senses that inform the organism regarding the homeostatic state of the body. Furthermore, the mentalization of the body involves not only perceptual integration and subsequent inferences, but also action and thus sensorimotor integration (active inference). Accordingly, we have made the radical claim that in early infancy when the motor system is immature, proximal interactions are necessary for the active mentalization of interoceptive states and therefore the corresponding core aspects of the minimal self. There is therefore a continuity between the minimal and the interactive, social self.

References

Adams, R. A., Shipp, S. & Friston, K.J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure & Function, 218,* 611-43.

Agren, G., Lundeberg, T., Uvnäs-Moberg, K. & Sato, A. (1995). The oxytocin antagonist 1-deamino-2-d-Tyr-(Oet)-4-Thr-8-Orn-oxytocin reverses the increase in the withdrawal response latency to thermal, but not mechanical nociceptive stimuli following oxytocin administration or massage-like stroking in rats. *Neuroscience Letters* 187, 49-52.

Ainley, V., Sel A & Tsakiris M (under review). Heartfelt Agency: Interoceptive Accuracy is Associated with Greater Intentional Binding.

Ainley, V., Apps, M.A.J., Fotopoulou, A., Tsakiris, M. (2016) 'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy. Phil. Trans. R. Soc. B 371: 20160003. http://dx.doi.org/10.1098/rstb.2016.0003Ainsworth, M. D. S., Blehar, M., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation. Hillside*, New Jersey: Erlbaum.

Anzieu, D. (1989). *The Skin Ego*. New Haven and London: Yale University Press. Trans. Chris Turner.

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews, 41*, 85–97.

Aron, L. (1998). *The clinical body and the reflexive mind*. In L. Aron & F. Sommer Anderson (eds.) Relational Perspectives on the Body. Hillsdale, NJ The Analytic Press 1998,

Baier, B., & Karnath, H. O. (2008). Tight link between our sense of limb ownership and self-awareness of actions. *Stroke*, 39, 486–488.

Bandura, A. (1967). The role of modeling processes in personality development. In W. W. Hartup & N. L. Smothergill (Eds.), *The young child*. Washington, DC: National Association for the Education of Young Children.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. Nature Reviews. *Neuroscience, 16*, 419–29.

Beebe, B., & Lachmann, F. M. (1988). Co-constructing inner and relational processes: Self and mutual regulation in infant research and adult treatment. *Psychoanalytic Psychology, 1*5,*1*5, 480-516

Bernhardt, B. C., & Singer, T. (2012). The neural basis of empathy. *Annual Review of Neuroscience, 35*, 1–23.

Besharati, S., Forkel, S.J., Kopelman, M., Solms, M., Jenkinson, & Fotopoulou, A. (2016) Mentalising the Body: Spatial and Social Cognition in Anosognosia for Hemiplegia, *Brain* , 139(Pt 3):971-85. Bird, G., & Viding, E. (2014). Neuroscience and Biobehavioral Reviews The self to other model of empathy : Providing a new framework for understanding empathy impairments in psychopathy , autism , and alexithymia. *Neuroscience and Biobehavioral Reviews, 47,* 520–532.

Blanke, O., and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Science*, *13*, 7–13.

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience, 13,,* 556–571

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron, 88,* 145-66.

Botvinick, M., & Cohen, J. (1998). Rubber hands “feel” touch that eyes see. Nature, (February), 1998–1998.

Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment.* New York: Basic.

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation: Anxiety and anger.* New York: Basic.

Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss: Sadness and depression.* New York: Basic.

Brener, J., Liu, X., & Ring, C. (1993). A method of constant stimuli for examining heartbeat detection: Comparison with the Brener–Kluvitse and Whitehead methods. *Psychophysiology*, *30*, 657–665.

Bretherton, I., & Munholland, K.A. (1999). Internal working models revisited. In J. Cassidy & P.R. Shaver (Eds.), Handbook of attachment: Theory, research, and clinical applications (pp. 89–111). New York: Guilford Press.

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive processing*, *14*(4), 411-427.

Carhart-Harris, R., & Friston, K. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain*, *133*, 1265–1283.

Cassidy, J., & Shaver, P. R. (Eds.). (2008). *Handbook of attachment: Theory, research, and clinical applications* (2nd edition). New York: Guilford Press.

Ciaunica, A. & Fotopoulou, A. (2016). *The Touched Self: Psychological and Philosophical Perspectives on Proximal Intersubjectivity and the Self*. Forthcoming in Durt C., Fuchs T., and Tewes C. (eds). Embodiment, Enaction, and CultureInvestigating the Constitution of the Shared World. Cambridge MA: MIT Press. In press.

Clark, A., (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences,* 36(3) 1-73.

Conger, J. P. (1994). *The body in recovery: Somatic psychotherapy and the self*. Frog Books.

Cowie D., Makin, T.R., & Bremner, A.J. (2013). Children’s responses to the rubber-hand illusion reveal dissociable pathways in body representation. *Psychological Science*, 24(5) 762–769.

Craig, A. D. (2009). How do you feel-now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10,* 59-70.

Craig, A., D. (2010). The sentient self. *Brain Structure and Function,* 214(5-6), 563–577.

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A. & Dolan, R. D. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience, 7,* 189-195195

Critchley, H. D., & Nagai, Y. (2012). How emotions are shaped by bodily states. *Emotion Review, 4,* 163–168.

Critchley, H. D., Eccles, J., & Garfinkel, S. N. (2013). Interaction between cognition, emotion, and the autonomic nervous system. *Handbook of Clinical Neurology, 117*, 59–77.

Critchley, H. D., & Harrison, N. a. N. (2013). Visceral Influences on Brain and Behavior. *Neuron, 77,* 624–638.

Crucianelli, L., Metcalf, N. K., Fotopoulou, A. K., & Jenkinson, P. M. (2013). Bodily pleasure matters: velocity of touch modulates body ownership during the rubber hand illusion. *Frontiers in Psychology, 4*, 703.

Crucianelli, L., Cardi, V., Treasure, J., Jenkinson, P., & Fotopoulou, A. (2016). The perception of affective touch in Anorexia Nervosa. *Psychiatry Research,* 239, 72-78

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: G.P. Putnam’s Sons.

Damasio, A. R. (2010). *Self comes to mind. Constructing the conscious brain*. London: Heineman.

Decety, J. (2011). The neuroevolution of empathy. Annals of the New York Academy of Sciences, 1231(1), 35–45.

Decety, J. & Fotopoulou, A. (2015). Why empathy has a beneficial impact on others in medicine: Unifying theories. *Frontiers in Behavioral Neuroscience* **8**. (doi:10.3389/fnbeh.2014.00457).

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., Cusack, R., Lawrence, A. D. & Dalgleish, T. (2010). Listening to your heart. How interoception shapes emotion experience and intuitive decision making. *Psychological Science, 21*,, 1835–1844. [20]

Esposito, G., Yoshida, S., Ohnishi, R., Tsuneoka, Y., del Carmen Rostagno, M., Yokota, S., Okabe, S., Kamiya, K., Hoshino, M., Shimizu, Masaki., Venuti, Paola., Kikusui, T., Kato, T. & Kuroda, K. O. (2013). Infant calming responses during maternal carrying in humans and mice. *Current Biology, 23,* 739-745.

Fairhurst, M. T., Löken, L., & Grossmann, T. (2014). Physiological and behavioral responses reveal 9-month-old infants' sensitivity to pleasant touch. *Psychological science*, *25*(5), 1124-1131.

Farmer, H. & Tsakiris, M. (2012). The bodily social self: A link between phenomenal and narrative selfhood. *Review of Philosophy and Psychology*. *3,* 125–144.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*, *4*, 215.

Feldman, R. & Eidelman, A. I. (2003). Skin-to-skin contact (Kangaroo Care) accelerates autonomic and neurobehavioural maturation in preterm infants. *Developmental Medicine & Child Neurology*, 4, 274-281.

Feldman, R. (2007). Infant biological foundations synchrony and developmental outcomes. *Current Directions in Psychological Science*, *16*, 340–345.

Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., & Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development, 34*, 569–577.

Filippetti, M. L., Johnson, M. H., Lloyd-Fox, S., Dragovic, D., & Farroni, T. (2013). Body perception in newborns. *Current Biology, 23*, 2413-2416.

Fletcher, P. & Fotopoulou, A. (2016). Abnormalities of agency. Oxford Textbook on Agency, Eds. Patrick Haggard and Baruch Eitam. Oxford: University Press. In press

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self*. London: Karnac..

Fonagy, P., Gergely, G & Target, M. (2007) The parent–infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry*, 48, 288-328.

Fonagy P. A. (2008). A genuinely developmental theory of sexual enjoyment and its implications for psychoanalytic technique. The *Journal of the American Psychoanalytic Association. 56*,11-36.*56*,11-36.

Fotopoulou, A. (2012a). Illusions and delusions in anosognosia for hemiplegia: from motor predictions to prior beliefs. *Brain*, *135*, 1344-1346.

Fotopoulou A. (2012b). Towards Psychodynamic Neuroscience. In A. Fotopoulou, M. Conway, & D. Pfaff. (Eds.) *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*. *Oxford University Press*.

Fotopoulou, A. (2013). Beyond the reward principle: Consciousness as precision seeking. *Neuropsychoanalysis*, *15*.

Fotopoulou, A. (2014). Time to get rid of the 'Modular' in neuropsychology: A unified theory of anosognosia as aberrant predictive coding. *Journal of Neuropsychology*, *8*, 1-19.

Fotopoulou, A. (2015). The virtual bodily self: Mentalisation of the body as revealed in anosognosia for hemiplegia. *Consciousness and Cognition*. 33, 500-10.

Fox, N. A., Schmidt, L. A., Henderson, H. A., & Marshall, P. J. (2007). Developmental Psychophysiology: Conceptual and Methodological Issues. *Handbook of Psychophysiology*, (1995), 453–481.

Fracasso, M. P., Porges, S. W., Lamb, M. E., & Rosenberg, A. A. (1994). Cardiac Activity in Infancy: Reliability and stability of individual differences.stability of individual differences. *Infant Behavior and Development*, *17*, 277–284.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London*, *360*, 815–36.

Friston, K. (2010). The free-energy principle: A unified brain theory. *Nature Reviews Neuroscience*, *11*, 127-138.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., et al (2012). Dopamine, affordance and active inference. PLoS Computational Biology, 8(1), e1002327.

Fuchs, T. & De Jaegher, H. (2009). Enactive Intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences, 8,* 465-486.

Gallace, A., & Spence, C. (2010). The science of interpersonal touch: an overview. *Neurosci Biobehav Rev, 34*(2), 246-259. doi: 10.1016/j.neubiorev.2008.10.004

Gallagher, S. & Meltzoff, A. (1996). The earliest sense of self and others: Merleau-Ponty and Recent Development Studies. *Philosophical Psychology 9*, 211-233.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science, *Trends in Cognitive Sciences 4*, 14–21.

Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies, 8*, 83–108.

Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology, 11*, 199–217.

Gallagher, S. (2005). *How the body shapes the mind*. USA: Oxford University Press.

Gallese, V. (2005). "Being like me": Self-other identity, mirror neurons and empathy. In S. Hurley & N. Chater (Eds.), Perspectives on imitation: From cognitive neuroscience to social science

(Vol. 1, pp. 101–118). Cambridge, MA: MIT Press.

Gallese V. (2013). Mirror neurons, embodied simulation and a second-person approach to mind reading. *Cortex 49*, 2954–2956.

Gallotti, M. & Frith C. (2013). Social cognition in the We-Mode. *Trends in Cognitive Sciences 17*, 160-5.

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., & Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology, 104*, 65–74.

Gentsch, A., Panagiotopoulou, E., & Fotopoulou, A. (2015). Active interpersonal touch gives rise to the social softness illusion. *Current Biology, 25,* 1-6.

Gentsch, A., Crucianelli, L., Jenkinson, P. & Fotopoulou, A. (2016). The touched self: affective touch and body awareness in health and disease. In H. Olausson, J. Wessberg, I. Morrison, F. McGlone (Eds.), *Affective Touch and the Neurophysiology of CT Afferents*. Springer. In press.

Gergely, G. & Watson, J. S. (1999). Early social-emotional development: Contingency perception and the social-biofeedback model. In P. Rochat (Ed.). *Early Socialization*. Mahwah, NJ: Lawrence Erlbaum Associates Inc

Gill, D.D. (1994). *Psychoanalysis in transition*. Hillsdale, NJ The Analytic Press.

Greenberg, J. (1991) *Oedipus and Beyond.* MA: Harvard University Press.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5,* 382–385.

Haggard, P., & Tsakiris, M. (2009). The Experience of Agency. *Current Directions in Psychological Science, 18,* 242–246.

Harlow, H. F. & Zimmermann, R. R. (1958). The development of affectional responses in infant monkeys. *Proceedings of the American Philosophical Society*. *102,* 501-509

Hart, S., Field, T., Hernandez-Reif, M., Nearing, G. , Shaw, S., Schanberg, S. & Kuhn, C. (2001) Anorexia Nervosa symptoms are reduced by massage therapy, *Eating Disorders, 9*, 289-299.

Herbert, B. M., & Pollatos, O. (2012). The body in the mind: On the relationship between interoception and embodiment.relationship between interoception and embodiment. *Topics in Cognitive Science, 4,* 692–704..

Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., & Jaskolka, A. R. (2006). Touch communicates distinct emotions. *Emotion, 6,* 528-533.

Henson, R. N. & Gagnepain, P. (2010) Predictive, interactive multiple memory systems. *Hippocampus 20*, 1315–1326.

Hill-Soderlund AL, Mills-Koonce WR, Propper C, Calkins SD, Granger DA, Moore GA, Gariepy JL, Cox MJ. (2008) Parasympathetic and sympathetic responses to the strange situation in infants and mothers from avoidant and securely attached dyads. *Dev Psychobiol. 50*(4):361-76. doi: 10.1002/dev.20302.

Hinton, G. E. & von Camp, D. (1993). Keeping neural networks simple by minimizing the description length of weights. In: Proceedings of COLT-93, 5–13

Hobson, R. P. & Lee A. (1999). Imitation and identification in autism. *Journal of Child Psychology & Psychiatry, 40*, 649-659.

Hurter, S., Paloyelis, Y., de C. Williams, A.C. & Fotopoulou, A. (2014). Partners' empathy increases pain ratings: Effects of perceived empathy and attachment style on pain report and display. *The Journal of Pain 15,* 934-944.

Insel, T.R. (2000). Toward a Neurobiology of Attachment. *Review of General Psychology 4*, 176-185.

Jenkinson, P.M. & Fotopoulou, A. (2014). Understanding Babiski's Anosognosia: 100 years later. *Cortex*, *61*, 1-4.

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Comput Biol*, *9*(6), e1003094.

Karnath, H.-O., & Baier, B. (2010). Right insula for our sense of limb ownership and self-awareness of actions. *Brain Structure and Function, 214*, 411–417.

Kehoe, P. & Blass, E. M. (1986). Behaviorally functional opioid systems in infant rats: II. Evidence for pharmacological, physiological, and psychological mediation of pain and stress. *Behavioral Neuroscience, 100,* 624.

Klinnert, M. D., Emde, R. N., Butterfield, P. & Campos, J. J. (1986). Social referencing: the infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology 22*, 427–432.

Korosi, A. & Baram, T.Z. 2010 Plasticity of the stress response early in life: mechanisms and significance. *Developmental psychobiology* 52, 661-670.

Krahe, C., Springer, A., Weinman, J. A., & Fotopoulou, A. (2013). The social modulation of pain: Others as predictive signals of salience – a systematic review. *Frontiers in Human Neuroscience, 7*, 386.

Krahé, C., Paloyelis, Y., Condon, H., Jenkinson, P. M., Williams, S. C. R., & Fotopoulou, A. (2015). Attachment style moderates partner presence effects on pain: A laser-evoked potentials study. *Social Cognitive and Affective Neuroscience, 10*, 1030-1037.

Krahé, C., Drabek, M., Paloyelis, Y., & Fotopoulou, A. (2016). Affective touch and attachement style modulate pain: A laser-evoked potentials study. *Philosophical Transactions of the Royal Society B.* 371 20160009; DOI: 10.1098/rstb.2016.0009. Published 10 October 2016

Legrand, D. (2006). The bodily self: The sensorimotor roots of pre-reflective self-consciousness. *Phenomenology and Cognitive Sciences, 5*, 89–118.

Limanowski, J., & Blankenburg, F. (2015). Network activity underlying the illusory self-attribution of a dummy arm. *Human Brain Mapping*, 36(6):2284-304

Lloyd, D. M., Gillis, V., Lewis, E., Farrell, M. J., & Morrison, I. (2013). Pleasant touch moderates the subjective but not objective aspects of body perception. *Frontiers Behavioral Neuroscience, 7*, 207.

Löken, L. S., Wessberg, J., Morrison, I., McGlone, F., & Olausson, H. (2009). Coding of pleasant touch by unmyelinated afferents in humans. *Nature Neuroscience, 12*, 547-548.

Lundeberg, T., Uvnäs-Moberg, K., Ågren, G. & Bruzelius, G. (1994). Anti-nociceptive effects of oxytocin in rats and mice. *Neuroscience Letters 170*, 153-157.

Mahler, S. and Pine, M.M. and F., Bergman, A. (1973). *The Psychological Birth of the Human Infant*. New York: Basic Books.

Maister, L., Tsiakkas, E., & Tsakiris, M. (2013). I Feel Your Fear: Shared Touch Between Faces Facilitates Recognition of Fearful Facial Expressions. Emotion, 13(1), 7-13

McLaughlin, K. A., Sheridan, M. A., Tibu, F., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2015). Causal effects of the early caregiving environment on development of stress response systems in children. *Proceedings of the National Academy of Sciences, 112*, 5637–5642.

McNally, G. P., Johansen, J. P. & Blair, H. T. (2011) Placing prediction into the fear circuit. *Trends Neuroscience, 34*, 283–292.

Meltzoff, A. (1989). Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology 25*, 954-962.

Meltzoff, A & Moore M. K. (2000). Imitation of facial and manual gestures by human neonates. *Infant Development: The essential readings*, Eds. D. Muir & A. Slater. London: Blackwell, pp. 167-175.

Meredith, P. (2013). A review of the evidence regarding associations between attachment theory and experimentally induced pain. *Current Pain and Headache Reports 17*, 1-9.

Merleau-Ponty, M. (1960), "Les relations avec autrui chez l'enfant", Paris, France: Cours de Sorbonne, trans. by W. Cobb, "The Child‟s Relations with Others", in: M. Merleau – Ponty (1964), *The Primacy of Perception And Other Essays on Phenomenological Psychology, the Philosophy of Art, History and Politics,* J. M. Edie (ed.), Evanston, IL: Northwestern University Press, pp. 96 – 155.

Merleau-Ponty, M. (1945). *Phénoménologie de la perception*, Paris: Éditions Gallimard; English translation: C. Smith. *Phenomenology of Perception*, London: Routledge and Kegan Paul, 1962.

Morrison, I., Bjornsdotter, M., & Olausson, H. (2011). Vicarious responses to social touch in posterior insular cortex are tuned to pleasant caressing speeds. *The Journal of Neuroscience, 31*, 9554-9562.

Morrison, I., Loken, L. S., & Olausson, H. (2010). The skin as a social organ. *Experimental Brain Research, 204*, 305-314.

Morton, J., & Frith, U. (1995). *Causal modeling: Structural approaches to developmental psychopathology***.** In D. Cicchetti & D. Cohen (Eds.), Developmental Psychopathology. (pp. 357-390). New York: Wiley.

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., & Spence, C. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 13169-13173.

Müller, U., Carpendale, J. I. M., Bibok, M. B., & Racine, T. P. (2006). Subjectivity, identification and differentiation: Key issues in early social development. In R. P. Hobson, G. Chidambi, A. Lee, & J. Meyer, Foundations for self-awareness: An exploration through autism. *Monographs of the Society for Research in Child Development, 71*.

Nelson, E.E. & Panksepp, J. (1998) Brain substrates of infant–mother attachment: contributions of opioids, oxytocin, and norepinephrine. *Neuroscience & Biobehavioral Reviews, 22*, 437-452.

Paladino, M. P., Mazzurega, M., Pavani, F., & Schubert, T. W. (2010). Synchronous multisensory stimulation blurs self-other boundaries. *Psychological Science*, 21(9), 1202-1207.

Paloyelis, Y., Doyle, O.M., Zelaya, F.O., Maltezos, S., Williams, S.O., Fotopoulou, A., & Howard, M.A. (2014). A spatiotemporal profile of in vivo cerebral blood flow changes following intranasal oxytocin in humans. *Biological Psychiatry, 79*, 693-705.

Paloyelis, Y., Krahé, C., Maltezos, S., Williams, S.C., Howard, M.A. & Fotopoulou, A. (2015) The analgesic effect of oxytocin in humans: A double-blinded placebo controlled cross-over study Using Laser-Evoked Potentials. *Journal of Neuroendocrinology, 28*, 1-10.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report of pubertal status: Reliability, validity, and initial norms. *Journal of Youth & Adolescence, 17,* 117–133.

Pezzulo, G., Rigoli, F. & Friston, K. (2015) Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology, 134,* 17-35.

Pollatos, O., Kurz, A. L., Albrecht, J., Schreder, T., Kleemann, A. M., Schöpf, V., Kopietz, R., Wiesmann, M. & Schandry, R. (2008). Reduced perception of bodily signals in anorexia nervosa. *Eating Behaviors, 9*, 381–388.

Poulin-Dubois, D., Brooker, I. & Chow, V. (2009). The developmental origins of naïve psychology in infancy. *Advances in Child Development and Behavior, 37,* 55-104.

Povinelli, D. J. & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. Jointly published in *Mind and Language, 19,* 1-28, and S. Hurley & M. Nudds (Eds.) *Rational Animals (2006)* Oxford University Press, Oxford.

Preston, S. D., & Hofelich, A. J. (2012). The many faces of empathy: parsing empathic phenomena through a proximate, dynamic-systems view of representing the other in the self. *Emotion Review, 4,* 24–33.

Rao, R.P., Ballard, D.H., (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience,* 2, 79–87.

Reddy, V. (2003). On being the object of attention: implications for self–other consciousness. Trends in Cognitive Sciences, 7(9), 397–402.

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

Rochat, P., & Morgan, R. (1995). Spatial determinants in the perception of selfproduced leg movements by 3- to 5-month-old infants. *Developmental Psychology, 31, 626-636.*

Ronchi, R., Bello-Ruiz, J., Lukowska, M., Herbelin, B., Cabrilo, I., Schaller, K., & Blanke, O. (2015). Right insular damage decreases heartbeat awareness and alters cardio-visual effects on bodily self-consciousness. *Neuropsychologia, 70,* 11–20.

Rothbart, M., Ziaie, H., & O'Boyle, C. (1992). Self-regulation and emotion in infancy. In N. Eisenberg, & R. Fabes (Eds.), *Emotion and Its Regulation in Early Development* (pp. 7-23). San Francisco: Jossey-Bass/Pfeiffer.

Sambo, C.F., Howard, M., Kopelman, M., Williams, S. & Fotopoulou, A. (2010). Knowing you care: Effects of perceived empathy and attachment style on pain perception. *Pain 151*, 687-693.

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology, 18*, 483–488.

Schauder, K. B., Mash, L. E., Bryant, L. K., & Cascio, C. J. (2015). Interoceptive ability and body awareness in autism spectrum disorder. *Journal of Experimental Child Psychology, 131,* 193–200.

Schechtman, M. (1996). *The Constitution of Selves.* Ithaca: Cornell University Press.

Scheler, M. (1913). *The Nature of Sympathy*. Trans. Peter Heath. *New York: Archon Books. 1970.*

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G. & Schlicht, T. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36,* 393–414.

Seth, A.K., Suzuki, K. & Critchley, H.D., (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology 2,* 1-16.

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Phil. Trans. R. Soc. B*, *371*(1708), 20160007.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Science*, *17,* 565-573.

Sforza, A., Bufalari, I., Haggard, P., Aglioti, S. M., (2010). My face in yours: Visuo-tactile facial stimulation influences sense of identity. *Social Neuroscience* 5.

Sharp, H., Pickles, A., Meaney, M., Marshall, K., Tibu, F. & Hill, J. (2012). Frequency of infant stroking reported by mothers moderates the effect of prenatal depression on infant behavioural and physiological outcomes. *PLoS One*, 7 (10)

Sharp, H., Pickles, A., Meaney, M., Marshall, K., Tibu, F., & Hill, J. (2012). Frequency of infant stroking reported by mothers moderates the effect of prenatal depression on infant behavioural and physiological outcomes. *PLoS One, 7,* e45446.

Slade, A. (2000). The development and organization of attachment: Implications for psychoanalysis. *The Journal of the American Psychoanalytic Association, 48,* 1147-1174.

Solms, M. (2013). The conscious Id. *Neuropsychoanalysis, 15,* 5-85.

Stern, D. (1985). *The Interpersonal World of the Infant,* New York: Basic Books.

Stevens, J., Schmied, V., Burns, E., & Dahlen, H. (2014). Immediate or early skin-to-skin contact after a Caesarean section: a review of the literature. *Maternal & child nutrition, 10*(4), 456-473.

Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia, 51*, 2909–2917.

Tajadura-Jiménez, A., Grehl, S., & Tsakiris, M. (2012). The other in me: interpersonal multisensory stimulation changes the mental representation of the self. *PloS one, 7*(7), e40682.

Tajadura-Jiménez, A., & Tsakiris, M. (2014). Balancing the "inner" and the "outer" self: Interoceptive sensitivity modulates self–other boundaries. *Journal of Experimental Psychology: General, 143*, 736–744.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences. 28*, 675–691.

Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., & Fink, G. R. (2007). Neural signatures of body ownership: A sensory network for bodily self-consciousness. *Cerebral Cortex, 17,* 2235–2244.

Tsakiris, M. (2008). Looking for myself: Current multisensory input alters self-face recognition. *PLoS ONE*, 3e4040.

Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia, 48*, 703–712.

Tsakiris M (2016) The multisensory basis of the self: From body to identity to others. *Quarterly Journal of Experimental Psychology, 17*:1-13

Tsakiris, M., Tajadura-Jiménez, A., & Costantini, M. (2011). Just a heartbeat away from one's body: interoceptive sensitivity predicts malleability of body-representations. *Proceedings. Biological Sciences / The Royal Society, 278*, 2470–2476.

Uvnäs-Moberg, K., Handlin, L. & Petersson, M. (2014). Self-soothing behaviors with particular reference to oxytocin release induced by non-noxious sensory stimulation. *Frontiers in Psychology 5*, 1529.

Vallbo, A. B., Olausson, H., & Wessberg, J. (1999). Unmyelinated afferents constitute a second system coding tactile stimuli of the human hairy skin. *Journal of Neurophysiology, 81*, 2753-2763.

Varela F., Thompson E. & Rosch E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA: MIT Press.

von Helmholtz, H. (1878/1971). The facts of perception. In: Kahl, R. (Ed.), Selected Writings of Herman von Helmholtz. Weslyan University Press.

Weaver, A., Richardson, R., Worlein, J., Waal, F.D. & Laudenslager, M. (2004). Response to social challenge in young bonnet (Macaca radiata) and pigtail (Macaca nemestrina) macaques is relate to early maternal experiences. *American journal of primatology*, 62, 243-259.

Weiss, W. J., Wilson, P. W., & Morrison, D. (2004). Maternal tactile stimulation and the neurodevelopment of low birth weight infants. *Infancy, 5*, 85–107.

Welsh, T. (2006). Do neonates display innate self-awareness? Why neonatal imitation fails to provide sufficient grounds for innate self- and other-Awareness. *Philosophical Psychology, 19*, 221-238.

Winnicot, D.W. (1972). Basis for self in body. *International Journal of Child Psychotherapy*, 1(1), 7-16.

Zahavi, D., (2005). *Subjectivity and Selfhood: Investigating the first-person perspective*, Cambridge, MA: The MIT Press.

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy and Shame*. Oxford: Oxford University Press.

Zahavi, D. (2015a). Self and other: From pure ego to co-constituted we. *Continental Philosophy Review, 48*, 143-160.

Zahavi, D. (2015b) You, me and we: The sharing of emotional experiences. *Journal of Consciousness Studies, 22,* 84-101.

Zahavi, D. (2016). Thin, thinner, thinnest: Defining the minimal self. Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World. red. / Christoph Durt; Thomas Fuchs; Christian Tewes. MIT Press, 2016.

Zahavi, D., & Rochat, P. (2015). Empathy≠Sharing: Perspectives from phenomenology and developmental psychology. *Consciousness and Cognition, 36*, 543-553.

Zeller, D., Litvak, V., Friston, K.J., Classen, J., 2014. Sensory processing and the rubber hand illusion – an evoked potentials study. J. Cogn. Neurosci. 27, 573–582.