Mentalizing homeostasis: the social origins of interoceptive inference – Replies to Commentaries

Aikaterini Fotopoulou[a*] and Manos Tsakiris[b,c]

[a] Clinical, Educational & Health Psychology (CEHP) Research Department, Division of Psychology & Language Sciences, University College London, 1-19 Torrington Place, London WC1E 7HJ, UK; b Lab of Action & Body, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, UK; [c] The Warburg Institute, School of Advanced Study, University of London, Woburn Square, WC1H 0AB, London, UK

*Corresponding author. Email: a.fotopoulou@ucl.ac.uk

Our target article ("Mentalising Homeostasis; The Social Origins of Interoceptive Inference" –henceforth MH for short) drew a relatively large and varied set of responses from the invited commentators. These are a source of both delight and challenge. The delight stems from the fact that our interdisciplinary proposal regarding the social origins of interoception is met with wonderfully constructive, specifying and expansive responses across different fields, thus building a wider basis for a long-overdue dialogue between the fields. The challenge stems from the fact that to do justice to the variety and depth of some of the commentaries would require at least another, long and overloaded, interdisciplinary article. As our response cannot be as long, we selected three central facets of our proposal and discuss them in relation to the points raised by the commentators. In all three instances, we particularly focus on the following question: Does our proposal call for a reconsideration of existing concepts and findings as we and some commentators suggest, or are the phenomena we have highlighted best described by existing conceptualisations in the fields involved, as some other commentators claim? We will start from the most radical aspect of our proposal, namely the cognitive and emotional implications of caregiving at the time of the infant's motor immaturity, before moving into the issue of embodied versus cognitive mentalisation. We will conclude with somewhat lesser issues of terminology regarding the nature-nurture distinction and the self-other distinction.

1. Interoceptive Regulation in the Age of Motor Immaturity.

In our article, we argue that given the infants' dependency on their caregivers, interoceptive inferences necessarily rely on other people's actions towards the infant. Because interoceptive inference is central to the formation of the self, it follows that other people's actions are central to the formation of the self. Somewhat to our surprise, only three commentaries (by Atzil & Feldman Barrett, by Bolis & Schillbach, and by Friston) actually address this central aspect of our proposal. What these commentators share is the fact that they are neuroscientists. By contrast, the remaining commentaries are psychological, psychoanalytic, and philosophical, and do not explicitly address this aspect of our proposal.

Indeed, this central tenet of our proposal is embedded in wider theoretical insights from biology, founded on the notion of homeostasis (Cannon, 1929), or the maintenance of a relative stability in one's physiological states despite ongoing internal and external changes (e.g. variabilities in metabolic energy levels and the availability of food). Moreover, as Atzil and Feldman Barrett appropriately emphasise, decades of work in neurophysiology has revealed that adaptive homeostatic control requires both some immutable reactions (e.g. basic

autonomic or motor reflexes), as well as long-term adaptation to a changing and unpredictable world (e.g. the so-called allostatic mechanisms, Sterling and Eyer, 1988). For instance, humans need to select between several possibilities for action that could fulfil different homeostatic imperatives at different spatiotemporal scales (e.g. delay eating if there is a chance of encountering more danger currently).

We believe that among the reasons why some commentators may not have engaged with this central claim of MH is the fact that concepts such as interoception, homeostasis, and allostasis, and particularly their potential relation to social relating, have been relatively neglected in fields such as developmental psychology and philosophy. For example, the concept of 'interoception' does not even appear in the commentary of Rochat and Zahavi, let alone its potential relation to motor immaturity. Instead, these authors try to direct our attention to assumed distinctions and interpretations of well-known exteroceptive and proprioceptive phenomena in their respective fields (e.g. the interpretation of proprioceptive-visual contingency detection, or self-touch as evidence for a sharp self-other distinction), suggesting we are somewhat confused or undecided about the nature-nurture division (see issue 3 below).

In addition, in fields such as psychoanalysis, the constructs of interoception, homeostasis and the immaturity of the motor system have become too 'physical' – old-fashioned, Freudian-like ideas, perhaps too mechanistic to account for psychic life, as elegantly discussed by Sandberg and Busch (their own equally elegant objections and additions to our proposal are addressed below). In the commentary by Fonagy and Cambell, for example, our proposal regarding interoception, and particularly the role of touch, is favourably considered in two explicit instances. Nevertheless, in both instances, the authors are quick to go 'beyond interoception' and introduce mentalistic concepts of intersubjective communication. In the first instance they emphasise 'meta-communication'; in their words, 'We use physical touch as a metacommunication. It is a primary communication that affirms selfhood, but looked at in terms of more advanced levels of functioning, embodied mentalization holds significance beyond interoception. Physical touch is an ostensive cue beyond the physically organizing characteristics described' (Fonagy &Cambell, 2017, p. XX). In the second instance they emphasize epistemic trust: "Through the meeting of physical needs, touch affirms the reality and validity of the infant body's needs. This is in part about interoception but it also has powerful implications for the developing mind's opening of epistemic trust…"

We have been both inspired and constrained by such ideas (see next point). Nevertheless, we feel the reductionistic, mechanistic level of analysis that stresses the physical aspects of intersubjective encounters is of paramount importance to mental life, as are more higher-order mentalistic concepts. The former, however, are increasingly absent in psychoanalysis, while the latter abound. Some philosophers and developmental psychologists engage with such physiological concepts, but proprioception and exteroception, particularly vision and discriminatory touch, seem far more frequently considered in these fields than interoceptive modalities. Having offered this general observation, we now turn to commentaries that have actually addressed our central claim regarding interoceptive regulation at the time of motor immaturity and its implications for mental life.

Atzil & Barrett (2017) put forward a strong case for a self-organising, predicting adult organism that needs to engage in homeostatic control, but also importantly in allostatic regulation. In their words, "the brain predicts physiological needs and attempts to meet them *before* they are required by the body (e.g., if your blood pressure changes after you stand, you

faint). The ongoing processes aimed to maintain physiological regulation through prediction and change are called *allostasis"* (p. [1 in proof], emphasis in the original]. From their point of view, the importance of early motor immaturity and social dependency is obvious. Infants are completely dependent on their caregivers for allostasis. For Atzil and Barrett this dependency not only ensures survival and shapes interoceptive awareness and the minimal self as in our proposal, but can be linked to all learning, including conceptual learning. In this sense, they stress not only the embodied aspects of intersubjectivity we have emphasised but also 'caregiving' as an allostatically-relevant concept that is learned in childhood. If we understand their position correctly, they argue that multimodal, predictive, probabilistic models in the brain primarily serve allostasis and they are formed on the basis of statistical associations between environmental conditions and bodily sensations. Thus, appropriate caregiving (mediated by parent theory of mind, see point 2 below) leads to statistical associations between one's bodily sensations and one's social environment, which in turn are used to create predictive cues for future allostatic changes. Their view reads like a very interesting, neuroscientific companion to the equally sophisticated, but more psychological commentary by Fonagy and Campbell.

In terms of our proposal, it is certainly true that we only covered allostasis in passing in MH ([page 41 in TA]), particularly as we were trying to emphasise what happens when allostatic regulation is not yet developed. It is also true that there is a degree of tension between active inference and classic, behaviourist perspectives on associative learning, which has been covered elsewhere (Pezzulo et al., 2015) and merely referred to in MH, as it escapes its scope. Beyond these wider theoretical considerations, however, it is clear that we agree with the constraints that Atzil and Barrett wish to emphasise. There is only one point in their commentary that we feel we cannot support, if of course we have understood them correctly. Atzil and Barrett seem to suggest that our proposal, perhaps, and more generally Friston's free energy framework, support a "nurture" explanation in the classic nurture-nature debate about mental/brain development ([pages 2-3 in TA]). Our proposal certainly aims to link concepts such as homeostasis and interoception that are usually suggested by proponents of 'nature' explanations with concepts of constitutive intersubjectivity that are typically used by supporters of 'nurture' explanations. However, as we further explain in point 3 below, our proposal, and in our view the free energy principle, actually do not support either pole, and instead question the validity and usefulness of sharp distinctions between nature and nurture. Predictions, generative models, and the process of embodied mentalisation have both inherited and empirically learned aspects, or in a different language, genetic and epigenetic aspects. This is the point contested by Zahavi and Rochat and we address it in point 3 below.

In their rich commentary, Bolis and Schilbach seem in agreement with our and other intermediate positions regarding both the 'self-other' distinction and the 'nature-nurture' distinction. They do however suggest that some of the processes of embodied mentalisation we propose could be further specified computationally. We completely agree, and beyond the current interdisciplinary work, we aim to progress in this direction in both theoretical, computational research and in modelling empirical work. In their own work, Bolis and Schilbach have suggested that at least three intertwined predictive loops allow the brain to continuously learn to anticipate the consequences of action or activity on itself, on the world, and on other people: 1) The 'inner loop,' which involves the brain re-describing its own representations to itself; 2) the 'perception-action' loop whereby the agent predicts the consequences of her actions on the world, and 3) the 'self-other' loop that links the agent with other agents, using the exact same prediction-based mechanisms as involved in the other two loops. We are unclear on the reasons why these and only these loops should be singled

out among all possible generative models, but we trust that the authors would cover this point too had they been given the space to do so. Nevertheless, the scope of their proposal is impressive to say the least, and we can certainly see the relevance of what they term 'collective behaviour' for some of the experiences described in MH. Our only addition is to emphasize that certain of the generative models assumed in the 'loops' of this theory may not mature developmentally at the same time, or with the same rhythm. Hence, the computational specificity that the authors call for could also attempt to predict and model such asymmetric development, particularly given what we know about cognitive and brain development (see also commentaries by Atzil and Barrett and by Friston).

Friston's commentary 'self-evidencing babies' reads predictably and unsurprisingly as a generous, rich, and highly advanced consideration of the themes we have tried to explore in interdisciplinary terms. Clearly, Friston sees and in fact pioneers the reductionistic value of searching for the fundamental, embodied principles that organise mental life. We find the casting of a baby as conforming to the imperative of maximising the evidence of its own existence as particularly meaningful and compatible with our own perspective. Moreover, Friston offers formal and clear concepts regarding our proposal of the consequences of caregiving at the age of motor immaturity. Most pertinently, he refers to the concepts of 'Closing the Action-Perception cycle' and 'Generalised Synchrony'. The former captures clearly our proposals regarding the build-up of generative, interoceptive models in early infancy. Even if inherited priors are sufficient to lead to the anticipation of certain aspects of homeostatic feelings such as pain, hunger and warmth, or even more general emotions such as fear and joy (see commentary by Solms), the infants cannot close the action-perception cycle without the caregiver. No matter what the genetic prior of a particular homeostatic imperative (in contrast with Solms's views, see also below), it is the caregiver's embodied intervention that allows the loop to be closed and hence predictive, cognitive models of interoceptive experience to be formed. Friston's formal consideration of interpersonal contingency, synchrony and matching (i.e. in relation to the mathematical construct of 'general synchrony' from dynamic systems theory) is also useful to advance our proposal towards the kind of computational modelling that Bolis and Schilbach suggest we should next engage with. Lastly, Friston rightly mentions past work on action observation and action understanding that considered the attenuation of both ascending and descending proprioceptive prediction errors as a prerequisite for intersubjective inter-actions with others (Kilner, Friston, & Frith, 2007). Similar suggestions about the need to attenuate self-representations during empathic interactions with others have been put forward by Bird & Viding (2014). Our proposal here is more specific to how ascending and descending interoception prediction errors and their attenuation are crucial not simply for bodily self-awareness (see Tsakiris, 2016 and Ainley et al, 2016 for reviews) but for our intersubjective relations with others. It is precisely in this extension of interoceptive predictive coding in intersubjectivity that the role of contextualising interoceptive evidence can be empirically tested.

2. **Embodied Mentalisation and the Caregiver's Mind.**

The commentaries by Atzil and Feldman Barrett, by Beebe, Steele and Steele, by Campbell and Fonagy, and by Sandberg and Busch stress not only the embodied aspects of

mentalisation, but also the importance of parental mental states during caregiving. For example, Atzil and Feldman Barrett write, "While Fotopoulou and Tsakiris hypothesize that the caretaker's body is sufficient for the infant developing a conceptual representation of 'self' (Fotopoulou & Tsakiris, 2017), we hypothesize that caretakers use their body and brain to regulate the infant's allostasis, and by that promoting the acquisition of concepts (Atzil et al.; Barrett, 2017). The role of parental higher-order social cognition in infant development is supported by research on post-partum depressed mothers…" , p. [__ in TA]. As we mentioned in MH, we do not wish to deny the importance of high-order social cognition in the mentalisation of the infants' body and self. We further agree with the long psychoanalytic tradition, as well as the more specific, empirical work of some of the commentators such as Fonagy and colleagues, Beebe, Steele and Steele on such dimensions of infant-parent interactions and their importance in mental life.

In our perspective and in our empirical work, we fully support the idea that parental, high-order social cognition and particularly the ability of parents to understand and relate to their children as mindful bodies is crucial for mental development. As Rochat and Zahavi rightly point out, we have not really emphasised such later aspects of mentalisation and self-other distinction in MH, because we felt this is covered in a plethora of other publications, including by all these commentators. Our emphasis on embodied mentalisation and embodied interaction does not wish to subtract anything from these insights. It does however wish to add to them by proposing a continuity between embodied and pre-reflective infant-parental interactions and interactions that entail higher-order cognition. We further feel that the former has received far less attention than the latter in many fields, including cognitive neuroscience and psychoanalysis.

Our paper therefore emphasises the embodied aspects of mentalisation. On a simple level, consistently with theories of embodied cognition, we do believe that all aspects of embodiment can influence the mind. For example, we believe that a very tall person will build a different relation to the world than a very short person. Thus, we see no tension is assuming that a very agile parent will influence their child in a different way that a very slow-moving individual. Nevertheless, what the commentators seem to suggest goes beyond this level of analysis. To borrow the elegant words of Sandberg and Busch, we have emphasised the pre-reflective aspects of parental mentalisation that are conveyed in parental physicality, rather than in explicit thoughts, decisions or even words. We certainly do not believe in a rigid split between such physicality and mental states (see also point 3 below), and a lot of our own empirical work has been on embodied and motor cognition. Indeed, it is from this standpoint that we understand the difference between higher-order social cognition and more embodied social cognition.

Furthermore, we feel the latter is also a crucial but neglected aspect of how parents contribute to the infant's minimal self. To use the example offered by Sandberg and Busch, a mother walking with her crying infant may be a mentally different mother than the one who sits with her infant, but we do not believe that this difference need be of the 'thinking' type, let alone the 'theory of mind' type. The difference may be restricted exactly in the part of the mind that is the mother's physicality. The 'how she walks or sits' with her infant is of course mental as

much as it is physical. However, it does not need to entail explicitly thinking of the infant's mind, taking its perspective, feeling empathy, etc. It may instead manifest first and foremost at the embodied cognition level, as an embodied reaction, and indeed expressed towards the infant in a mother's soothing, calm, or high-strung, hyperactive physicality. Frequently of course the two levels of cognition convey similar things and we agree that there can be a tight interrelation between the two. Nevertheless, this similarity cannot be taken for granted in all cases and it does not diminish the importance of either pole. A father with a certain degree of alexithymia and obsessive compulsive tendencies, for example, may think and do very different things towards a crying infant. Both reactions may relate to this father's theory of mind, but their consequences towards the infant need not be the same. A physically-absent father and a mentally-absent father may share some psychic commonalities and be both damaging for their infants, but the effects may not be identical. Focusing on both embodied and higher-order cognitive aspects of mentalisation has the advantage of capturing the complexity involved in understanding how other people contribute to the formation of the minimal self and development more general.

3. **Questions of terminology, taxonomy and tautology.**

In their commentary, Rochat and Zahavi first discuss developmental considerations in relation to MH and then outline philosophical considerations. Both, however, seem to center on the self-other distinction and the nature-nurture distinction, while the philosophical part also touches on the mind-body distinction. Clearly, we cannot do justice to all these huge topics and their epistemological and ontological dimensions in this response. Rochat and Zahavi seem to take the view that the minimal self, what they term 'what-it-is-like-*for-me*-ness', is innate and distinct from the outset, in direct contrast to Azil and Feldman Barrett who take a social constructivist approach to mental life. These kind of debates are of course as old as some of these fields. From the standpoint of Rochat and Zahavi, our intermediate position in relation to the self-other and the nature-nurture distinction is 'ambiguous' at best, and they urge us to clarify it in this respect. From their point of view it seems that clarity on this topic is equal to choosing between two extremes, the minimal self is either innate or it is learned. Particularly in the philosophical part, no other alternative seems possible.

From the perspective of MH however, the minimal self is neither innate, nor constructed in development. In our assumptions, we follow many other scientists and scholars who feel that a simple distinction between nature and nurture is rather simplistic when it comes to the development of certain properties of the mind in health and in pathology. Specifically, we assume that humans, like presumably other mammals, are capable of developing a minimal self, i.e. they have the phylogenetically developed ability to develop one in ontogenesis given certain biological and social conditions. One important condition is the regulation of their homeostasis in infancy. This regulation in infancy is determined by embodied interactions with caregivers. However, Zahavi and Rochat further claim that it is not clear whether we are proposing that embodied interactions constitute the minimal self, or 'simply affect' its specific qualitative character. We are rather intrigued by their detailed quoting of our text in this respect, as we have proposed that in our view the minimal self is synonymous to the feeling qualities associated with being an embodied subject. Therefore, we believe that

embodied interactions indeed ‘shape’ these qualities and this shaping is synonymous to saying that they constitute the minimal self, or at least part of it (the rest may be subject to other nature-nurture couplings). In other terms, we are not sure that a distinction between giving rise to the ‘existence’ of a quality and giving rise to the particular ‘quality’ of a quality is very useful. We do of course understand that these commentators may be driven by concerns of ontology, but as we have explained in MH, given that without caregiving, infants die, we are less concerned with zombie-like states and more concerned with the mechanisms that lead individuals to develop their particular minimal self. Thus, in sum, we believe that experientiality and phenomenality are constituted both in phylogenetic and ontogenetic development and embodied interactions relate to the latter. In this sense, at least some parts of the minimal self are interpersonally constituted and embodied interactions shape the minimal self in each individual.

Finally, we are able to discuss only two of the many comments regarding the mind-body problem. Apparently, Rochat and Zahavi suggest that we attempted to explain the hard problem of consciousness because we have suggested that certain ‘subpersonal’ processes such as multisensory integration can give rise to phenomenal consciousness. While we do not think that phenomenal consciousness should be reduced epistemologically to neural processes, we indeed believe that conscious processes are also neural processes. We are also intrigued by the offered interpretation of evidence regarding ‘in-utero’ learning of ‘avoidance’ behaviours and preference for the smell of one’s own mother’s amniotic fluid as evidence in favour of an innate self and against the role of caregiving in embodied mentalisation. This sounds perhaps like the reverse argument of the psychoanalytic insistence that embodied interpersonal exchanges cannot be conceived without reference to higher-order mentalisation. In this instance, according to Rochat and Zahavi, the interaction of the infant’s body and the mother’s body does not really amount to social development until the infant is technically out of the mother’s body. It seems that before that moment, all learning, including the sensory learning of the mother’s body itself, should be seen as exclusively innate and unrelated to ‘proper intersubjective interactions’. We simply disagree: the interaction of the fetus with the environment of the mother’s womb, that would reflect a wide range of the mother’s mental and physiological states, could be understood as a basic form of bodily intersubjective exchanges. More generally, we think in light of what we have said before, it suffices to say that we disagree with such rigid distinctions between nature-nurture, self-other and mind-body.

Two briefer terminological qualifications are left. We entirely agree with Solms’s many clear and engaging points, including his claim regarding the difference between objects of homeostatic feelings, that relate less to other people, and those of emotional systems such as anger, or separation anxiety, that are primarily about other people. However, our proposal is not based on the objects that will satisfy a drive or instinct (in his terminology). We are not claiming that infants are hungry for love, or thirsty for attention. However, we are claiming that other people, and particularly their embodied caregiving, are necessary for allowing any object, or any change upon the state of the infant’s environment, to satisfy and regulate their homeostatic needs. Hence, it is not a matter of whether the infant can eat itself or another. It is a matter of whether it can eat anything without the other, or, whether it can thermoregulate

without the other. In fact, as Solms argued, the infant cannot afford trial and error when it comes to jumping off cliffs, or putting hands into plugs or choking. Yet, without a caregiver the infants actually risk all of these dangers. This is not only a truism in development, it is a matter of life and death for the infant.

Finally, we would like to thank Sandberg and Busch for their thoughtful suggestions regarding the role of illusions in mental life. We sincerely wish we could have the space to do justice to this point about the rubber hand illusion and the mental learning of 'reality'. In fact, in the previous theoretical work of one of us (e.g. Fotopoulou, 2015) and in current, ongoing empirical and theoretical work, we have indeed increasingly taken this exact perspective, inspired among others by Winnicot. Specifically, we envision the role of an emotional, second person perspective in creating temporary illusions of unity upon which the distinction between first and third person perspectives on embodiment can take place. Sadly, we cannot unpack these ideas here but we will certainly be revisiting them in the near future, and we hope to have the opportunity to seek the enriching perspective of these commentators again. We only wish to disagree in one point: we do not think this topic portrays a genuine challenge between science and psychoanalysis but rather a unique opportunity for dialogue and knowledge exchange.

References

Ainley V, Apps MAJ, Fotopoulou A & Tsakiris M (2016) Bodily Precision': A Predictive Coding Account of Individual Differences in Interoceptive Accuracy. Philosophical Transactions, Series B, Biological Sciences, 371(1708). pii: 20160003. doi: 10.1098/rstb.2016.0003.

Bird G, Viding E. (2014) The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. Neurosci Biobehav Rev.; 47:520-32.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. Cogn Process., 8(3), 159-166.

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

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