The role of interoceptive inference in theory of mind

Sasha Ondobaka *, James Kilner, Karl Friston

Wellcome Trust Centre for Neuroimaging, University College London, United Kingdom
Sobell Department for Motor Neuroscience and Movement Disorders, University College London, United Kingdom

ABSTRACT

Inferring the intentions and beliefs of another is an ability that is fundamental for social and affiliative interactions. A substantial amount of empirical evidence suggests that making sense of another's intentional and belief states (i.e. theory of mind) relies on exteroceptive (e.g. visual and auditory) and proprioceptive (i.e. motor) signals. Yet, despite its pivotal role in the guidance of behaviour, the role of the observer's interoceptive (visceral) processing in understanding another's internal states remains unexplored. Predicting and keeping track of interoceptive bodily states – which inform intentions and beliefs that guide behaviour – is one of the fundamental purposes of the human brain. In this paper, we will focus on the role of interoceptive predictions, prescribed by the free energy principle, in making sense of internal states that cause another's behaviour. We will discuss how multimodal expectations induced at deep (high) hierarchical levels – that necessarily entail interoceptive predictions – contribute to inference about others that is at the heart of theory of mind.

1. Understanding others' minds

1.1. Mechanisms for inferring others' minds

Understanding or inferring of another's intentions, feelings and beliefs is a hallmark of human social cognition often referred to as mentalising or having a Theory of Mind (ToM; Frith & Frith, 1999; Saxe & Kanwisher, 2003). ToM has been described as a cognitive ability to infer the mental states (intentions and beliefs) of others, through processing of their physical appearance and overt behaviour (e.g. clothes, bodily and facial expressions). Typically, neuronal computations underlying ToM have been associated with multimodal brain regions like the superior temporal sulcus (STS), temporoparietal junction (TPJ) and medial frontal cortex (MFC; Frith & Frith, 2006; Saxe, 2006). However, the nature of processing in higher multimodal regions that accumulates information from different streams remains poorly understood.

1.2. Predictive mechanisms in theory of mind

Recent views propose that predictive mechanisms could play a role in ToM (Hohwy, 2013; Kilner & Frith, 2008; Koster-Hale & Saxe, 2013). In brief, hypotheses about the intentions of others are tested against their observed behaviour by generating top-down predictions of that behaviour – and updating competing hypotheses on the basis of ensuing prediction error. Crucially, the repertoire of hypotheses that can be entertained is borrowed from the constructs (hypotheses) that cause one's own behaviour. This provides a nice explanation for the role of the "mirroring mechanism" (Rizzolatti & Sinigaglia, 2010) in action observation and theory of mind – and the integration of multimodal data inherent in descending multimodal predictions (Kilner & Frith, 2008; Ondobaka & Bekkering, 2013; Ondobaka, de Lange, Wittmann, Frith, & Bekkering, 2014).

However, this perspective only addresses the predicted consequences of movement and says little about the predicted consequences of internal bodily states that contextualise behaviour. In other words, it is unclear how processing of internal visceral/autonomic information (interoception) could contribute to the understanding of others' intentions. There are two ways of thinking about the role of interoception in ToM. The first relates to how processing of exteroceptive information about another's interoceptive state helps us to infer states of mind that cause their behaviour. The second rests on how knowing the interoceptive causes of our own behaviour helps us predict and infer another's. In this paper, we consider interoceptive inference as a special case of active inference, under the free energy principle (Friston, 2010) – and emphasise its potentially fundamental role in grounding the process of inferring another's state of mind from their perceived (motor and autonomic) behaviour (Fig. 1).

* Corresponding author at: Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom.
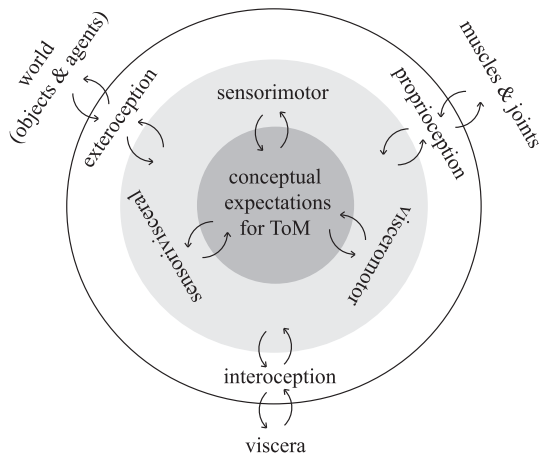   *E-mail address:* s.ondobaka@ucl.ac.uk (S. Ondobaka).

Fig. 1. Schematic representation of a hierarchical predictive neural model for ToM that includes interoception, exteroception and proprioception. White-to-dark grey colour scale represents the neural hierarchy, in which conceptual expectations (dark grey) that include interoception sit high (deep) in the hierarchy. Arrows indicate hierarchical message passing in forward and backward directions carrying prediction error and expectation/prediction, respectively.

2. The free energy and interoceptive inference

2.1. The free energy principle

The free energy principle requires the brain to generate continuous predictions in order to achieve its goal of minimizing free energy – an information theoretical quantity that reflects surprise or prediction error (Friston, 2009). Prediction errors are the difference between sensations and predictions of those sensations based upon an internal or generative model. The basic idea is that the brain constitutes a hierarchical generative model, through which the wealth of incoming information from the viscera, musculoskeletal system and the outside world is interpreted. This interpretation corresponds to inferring the causes of sensations in terms of representations or expectations that would generate the same sensory information, under the hierarchical model. Inferring the causes of visceral/autonomic, motor and sensory (e.g. visual, auditory) information corresponds to interoception, proprioception and exteroception, respectively. It is assumed that all the sensory streams in the brain are organised in a hierarchical fashion, with areas that sit higher (deeper) in the hierarchy representing more abstract information and generating expectations of lower levels (Ondobaka & Bekkering, 2013). At the apex of unimodal sensory hierarchies, multimodal brain regions encode conceptual expectations that necessarily accumulate unimodal information – or prediction errors (Fig. 1; Barrett, 2014; Ondobaka et al., 2014).

2.2. Two modes of free energy minimisation

Minimisation of free energy (prediction error) can occur in two distinct modes, we can either change our predictions to match sensations or we sample sensations to match predictions. Predictive coding typically refers to changing our predictions to match sensations (Friston, 2005; Rao & Ballard, 1999). Controlling motor and autonomic (visceral) system to experience sensations that match our predictions is known as active inference (Friston, Mattout, & Kilner, 2011; Joffily & Coricelli, 2013). These complementary modes of minimising (exteroceptive and proprioceptive/interoceptive) prediction errors correspond to what we generally view as perception or inference and action or motor/autonomic control respectively. It is crucial to note here that two modes of free energy minimisation exist in both proprioceptive and interoceptive domains.

Perception minimises free energy by concurrent dynamical updating of expectations about the causes of external (exteroceptive) and internal (interoceptive and proprioceptive) sensory inputs. For example, perception of a surprising object is associated with an attempt to suppress visual prediction error (Rao & Ballard, 1999). Action, on the other hand, minimises prediction error by directly altering sensory inputs through movement and visceral control that fulfil proprioceptive and interoceptive predictions. For example, movement of the arm is driven by classical motor reflects arcs in the spinal-cord to suppress proprioceptive prediction error – such that descending proprioceptive predictions become motor commands that are reflexively executed by striated muscles (Adams, Shipp, & Friston, 2013). Similarly, the intensity and frequency of on-going contractions of the heart muscle can be modulated to suppress the interoceptive prediction error signalling surprising interoceptive states related to e.g. blood pressure (e.g. Kumagai et al., 2012). This reflexive suppression of interoceptive prediction error corresponds to autonomic reflexes mediated by smooth muscles.

2.3. Free energy and experience of intention and emotion

Proprioceptive and interoceptive prediction errors (free energy) used in motor and autonomic control might relate to our experience of intention and emotion (Seth, 2013; Shipp, Adams, & Friston, 2013). Recent accounts of interoceptive inference have proposed that emotion could be understood from the perspective of hierarchical interoceptive inference (Joffily & Coricelli, 2013; Seth, 2013). For example, Seth (2013) views emotional content as the product of active inference about the likely internal and external causes of visceral changes. Joffily and Coricelli (2013) associated the rate of change of interoceptive prediction error with emotional valence, such that a shift from less expected/valued state (i.e. high free energy) to a more expected/valued state (i.e. low free energy) leads to positive valence. Conversely, negative valance corresponds to a shift from a low free energy to a high free energy state. Although the primary drive for motor and autonomic (visceral) control are descending proprioceptive and interoceptive predictions, these predictions are contextualised by deep hierarchical models that are necessarily accountable to conceptual representations that also generate exteroceptive predictions (Barrett, 2014; Barrett & Simmons, 2015). This enables the integration of exteroceptive information to contextualise (adaptive) motor and visceral responses. For example, when my interoceptive prediction errors signal hunger, I intend (expect) to move my arm to open the fridge because this is what I normally do when I feel (infer myself to be) hungry and find myself in the kitchen. We assume that these deep multimodal levels of representation that guide own behaviour play an important role in inference of others' intentions and emotion.

3. Inferring another's intentions and emotions

3.1. Interoceptive inference mechanism

Interoception refers to one of the fundamental purposes of the human brain – to predict and maintain internal bodily states within physiological bounds and relatively constant over time (Craig, 2009; Critchley & Harrison, 2013). Interoception or interoceptive inference can be viewed as a generalisation of active inference (Clark, 2013; Friston, 2010) to the processing of interoceptive signals carrying information about visceral states (e.g. heart rate, blood pressure, temperature). For example, recent computational work by Joffily and Coricelli (2013) suggests that rate of change of interoceptive free energy corresponds to experience of emotional valence and dynamical adaptation of behaviour. The interoceptive (visceral) processing hierarchy comprises of the brainstem

nuclei (e.g., nucleus of the solitary tract, parabrachial nucleus, and the central grey) carrying afferent inputs to the thalamus, hypothalamus, the amygdala and the neocortex (Craig, 2009; Kumagai et al., 2012). Cortical regions associated with interoceptive processing include the cingulate gyrus, insula, somatosensory cortex, medial and orbital frontal cortex. We assume that interoceptive information encoded by mechanical, chemical and thermal receptors is processed in the light of top-down predictions generated by neuronally encoded expectations at each level of the interoceptive hierarchy (Friston, 2010). Mechanistically, the recurrent message passing between different levels in the interoceptive hierarchy will aim to minimise prediction errors at each level. This is exactly the same predictive coding scheme proposed for predictive processing of exteroceptive (Friston, 2005; Rao & Ballard, 1999) and proprioceptive (Adams et al., 2013; Shipp et al., 2013) information and applied to mechanisms that underlie understanding of intentions in others (Kilner, Friston, & Frith, 2007; Ondobaka & Bekkering, 2013).

3.2. Interoceptive inference in ToM

In this paper, we suggest that delineating the interplay between interoceptive, proprioceptive and exteroceptive sensory processing is pivotal to understanding mechanisms that underlie ToM. We hypothesise that interoception, formulated under active inference, plays a fundamental role in ToM. From the active inference perspective, knowing the contents of another's mind can be demystified and simply recast as an optimal explanation for perceived (motor and visceral) behaviour in others – that would have been produced by ourselves, have we had been in the same intentional and emotional state.

Note that deep generative models permit inferences at multiple levels. For example, inferring the interoceptive (visceral) states of another necessarily constrains the hypothesis space of plausible explanations for their current behaviour (e.g. she wants to go back indoors because she's cold). Observer's predictions of their own interoceptive states that cause a feeling of being cold play a principal role in understanding this feeling when seeing someone shiver. This is implied because interoceptive predictions are generated in the observer's deep hierarchical model that concurrently generates exteroceptive predictions of seeing someone shiver. In this instance, the observation of shivering may induce an interoceptive or emotional contagion and empathy – implying that an observer can also be sympathetic to another's desires and intentions (e.g. to go back into the warmth). Clearly, this form of deep vicarious active inference requires the attenuation of (proprioceptive and interoceptive) prediction errors that would otherwise cause echopraxia or overt emotional contagion (Hess & Fischer, 2014; Heyes, 2011).

3.3. Contrasting interoception to exteroception and proprioception

To understand the fundamental role of interoception in ToM, we might consider its unique contribution to (social) cognition by contrasting it to proprioceptive and exteroceptive inference. First, compared to proprioception and exteroception, interoceptive sensations have a low degree of spatiotemporal acuity and do not typically reach conscious awareness. For example, it is difficult to localise a stomach ache in space and time. Similarly, there is little conscious access to the peristaltic contractions of the transverse colon or status of renal function. These low levels of acuity or resolution are a direct consequence of the generative models we have inherited to predict the continuously changing interoceptive signals reported by a large variety of receptors. Several interesting predictions arise from formal or structural differences in the generative models for interoception, relative to visual or tactile cues.

First, it suggests that we would be unable to integrate localised and transient cues of an interoceptive nature into beliefs about the emotional intentions of another. For example, a brief (one second) blush localised to one side of the face may be perceived as a brief change in ambient illumination, as opposed to a change in emotional state. Second, due to a high degree of exteroceptive–proprioceptive (i.e. visuomotor) correspondence, the effects of observed agent's actions can be directly mapped onto the observer's proprioceptive system (Rizzolatti & Sinigaglia, 2010). For example, both observer and the observed agent, can directly access exteroceptive information about the precise motor trajectories aimed at objects in the shared environment (Bach, Peelen, & Tipper, 2010). This exteroceptive information can then be combined with observer's proprioceptive models to infer movement intentions (Ondobaka et al., 2014).

In contrast, interoceptive–exteroceptive correspondence is rather poor, making the direct mapping of others' interoceptive states to our own interoception more challenging. In other words, there are far fewer interoceptive cues that can be observed (through exteroception) and consequently a much greater reliance upon counterfactual inference afforded by priors in generative models of our own interoceptive states. Similar to already proposed proprioceptive counterfactuals (Friston, Adams, Perrinet, & Breakspear, 2012; Seth, 2015), interoceptive counterfactuals or simulations can be viewed as expectations or hypotheses about future sensory states that are conditioned upon certain interoceptive states. Interoceptive counterfactuals rest upon deep hierarchical models that extend into the future and entail both exteroceptive and proprioceptive expectations.

3.3.1. From interoceptive inference to ToM

Whereas majority of the investigations on intentional and emotional aspects of ToM focus on exteroceptive and proprioceptive processing, the role of interoceptive processing has been somewhat neglected. Interoceptive states are informing and contextualizing behaviour by biasing perception and action towards fulfilling organism's physiological needs. At a low level of the hierarchy this corresponds to homeostasis, while higher levels provide contextual guidance for allostasis (Barrett, 2013; Gu & FitzGerald, 2014). In a similar vein, Seth (2013) has proposed a fundamental role for interoception in body ownership, emotion and selfhood. Crucially, in exactly the same way that the proprioceptive mirroring mechanism may reflect a re-purposing of generative models for inferring and causing one's own motor behaviour to infer another's, it has also been proposed that interoceptive inference can also be deployed to infer the interoceptive states of others (particularly in theoretical treatments of autism: e.g., Quattrocki & Friston, 2014). In other words, in the same way that we simulate how we would move given a particular intention – to explain how someone else is moving, we may simulate our bodily states corresponding to a particular feeling to understand how someone else is feeling.

While interoceptive inference has also been associated with processing of other individuals' emotional and feeling states (Singer et al., 2004; Wicker et al., 2003), the predictive mechanisms that integrate interoceptive information with visual and proprioceptive processing streams to model or read others' minds remain unclear. Recent work provides behavioural evidence for the role of interoception in social inference by showing that ability to recognise others' emotion from facial expression is correlated with perceivers' ability to report their own interoceptive states (Cook, Brewer, Shah, & Bird, 2013). This mismatch between precise information about another's motor behaviour, relative to their autonomic responses, speaks to the important role of (Bayesian) inference in reconciling multimodal cues to make inferences. It suggests that we will weight evidence about motor behaviour over

evidence for changes in autonomic status. This differential weighting suggests a useful perspective on theory of mind, from the perspective of predictive multisensory integration. For example, what would be the effect of incongruent cues from another's motor and autonomic responses (i.e., a gentle affiliative touch accompanied by piloerection and pupillary dilation)? Would these lead to a sense of uncertainty about the actor's real intent or would the motor behaviour predominate? In principle, these questions can be answered empirically using behavioural or neuroimaging techniques.

3.3.2. Differences between inferring mental and physical states

We have framed inference about the emotional states of others as a form of interoceptive mirroring mechanism (Singer et al., 2004; Wicker et al., 2003); however, there are clear differences between proprioceptive and interoceptive inference. A key difference is that the motor behaviour of another is readily accessible through visual and other exteroceptive cues. This is not necessarily the case for interoception; in the sense that many bodily states are hidden from (sensory) view. One might therefore ask is there are fundamental differences between inferring the behaviour of any inanimate object and inferring the internal state of another person. For motor behaviour, this is an easy question to answer because one can use one's own intentional expectations and sensorimotor predictions to furnish a simple explanation for the behaviour of another (provided one applies a suitable perspective taking transformation to visual cues). Same proprioceptive generative models are used in inference of external physical events (Schubotz, 2007). In other words, instead of having to learn a generative model of another's goal-directed behaviour, it is only necessary to repurpose one's own generative models that are acquired during early neural development.

However, for interoceptive cues (in the exteroceptive modality), this perspective taking is more difficult and sometimes perhaps impossible. For example, we cannot see our pupils dilate. This suggests that the emotional and intentional theory of mind has to be learned through interpersonal interactions, probably at an early stage of development, in which attachments are made. However, if one can learn to explain or associate cues about the interoceptive state of another in terms of one's own interoceptive state, all the deep hierarchical contingencies that underlie allostatic behaviour become available to predict what the other person may do next – and why they are doing it. It is in this sense that there may be something special about how we are particularly adept at inferring the drives and affiliative imperatives that contextualise interoception.

3.3.3. Conceptual (multimodal) inference in ToM

During social inference, ToM or mentalising regions combine interoceptive information with proprioceptive and exteroceptive signals (Brass, Ruby, & Spengler, 2009; Hayden, Nair, McCoy, & Platt, 2008; Pearson, Heilbronner, Barack, Hayden, & Platt, 2011). Whereas the effects of one's interoceptive states play a crucial role in behaviour (Critchley & Harrison, 2013), there is no direct one-to-one mapping of interoception onto proprioceptive and exteroceptive states. Interoceptive states are rather mapped, with exteroceptive and proprioceptive states onto multimodal constructs in a hierarchical fashion (Fig. 1). This means, each modality contextualises the others, through ascending prediction errors and resulting updates at deep or higher conceptual levels of the hierarchical model. For example, to maintain an expected heart rate, many (probabilistic) mappings to exteroceptive states (objects) and proprioceptive states (movements) could exist. Similarly, no one-to-one mapping exists between the agent's and the observer's interoceptive states. This pleiotropic (producing more than one effect) mapping precludes exact inference about another's interoceptive

states; e.g., their heart rate. Empirically speaking this form of vicarious (interoceptive) inference suggests that neuronal responses to changes in interoceptive cues should depend upon exteroceptive cues at and only at higher/conceptual levels of the hierarchy (e.g., cingulate, insular or prefrontal cortex). On the other hand, the fundamental advantage, offered by the probabilistic and pleiotropic mappings, is the potential to engage in simulation or counterfactual inference (Seth, 2015). Pleiotropic and probabilistic mappings between interoceptive and exteroceptive/proprioceptive states comfortably accommodate counterfactual conceptual inference – a central component of ToM. In principle, this sort of prediction can be tested using factorial neuroimaging experiments in which exteroceptive and interoceptive cues are manipulated orthogonally to elicit interactions in multimodal areas that generate top-down interoceptive predictions.

4. Conclusion

In summary, the inference about or understanding of models that cause another's behaviour in mentalising or ToM may be mediated by joint minimisation of hierarchical prediction errors (free energy) elicited by unexpected sensations. Multimodal expectations induced at deep (high) hierarchical levels – that necessarily entail interoceptive predictions – should play a fundamental role in inferring causes of sensory impressions produced by the behaviour of other people. Most current social cognition models suggest that exteroceptive (e.g. visual and auditory) and proprioceptive (i.e. motor) processing underlie inference about intentional and emotional states. Embedding interoception in a multimodal active inference framework may offer a more complete and plausible model of ToM. This extension may also provide a principled account of ToM that appeals to a sense of agency and selfhood – and that is equipped with the interoceptive aspects of affect and feeling states.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function, 218*(3), 611–643.

Bach, P., Peelen, M. V., & Tipper, S. P. (2010). On the role of object information in action observation: An fMRI study. *Cerebral Cortex, 20*(12), 2798–2809.

Barrett, L. F. (2013). Psychological construction: The Darwinian approach to the science of emotion. *Emotion Review, 5*(4), 379–389.

Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review, 6*(4), 292–297.

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

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364*(1528), 2359–2367.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181–204.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science, 24*(5), 723–732.

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

Critchley, H. D., & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron, 77*(4), 624–638.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836.

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13*(7), 293–301.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology, 3*.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics, 104*(1–2), 137–160.

Frith, C. D., & Frith, U. (1999). Interacting minds – A biological basis. *Science, 286* (5445), 1692–1695.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531–534.

Gu, X., & FitzGerald, T. H. (2014). Interoceptive inference: Homeostasis and decision-making. *Trends in Cognitive Sciences, 18*(6), 269–270.

Hayden, B. Y., Nair, A. C., McCoy, A. N., & Platt, M. L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron, 60*(1), 19–25.

Hess, U., & Fischer, A. (2014). Emotional mimicry: Why and when we mimic emotions. *Social and Personality Psychology Compass, 8*(2), 45–57.

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463.

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

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

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8*(3), 159–166.

Kilner, J. M., & Frith, C. D. (2008). Action observation: Inferring intentions without mirror neurons. *Current Biology, 18*(1), R32–R33.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron, 79*(5), 836–848.

Kumagai, H., Oshima, N., Matsuura, T., Iigaya, K., Imai, M., Onimaru, H., et al. (2012). Importance of rostral ventrolateral medulla neurons in determining efferent sympathetic nerve activity and blood pressure. *Hypertension Research, 35*(2), 132–141.

Ondobaka, S., & Bekkering, H. (2013). Conceptual and perceptuo-motor action control and action recognition. *Cortex, 49*(10), 2966–2967.

Ondobaka, S., de Lange, F. P., Wittmann, M., Frith, C. D., & Bekkering, H. (2014). Interplay between conceptual expectations and movement predictions underlies action understanding. *Cerebral Cortex*. bhu056.

Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., & Platt, M. L. (2011). Posterior cingulate cortex: Adapting behavior to a changing world. *Trends in Cognitive Sciences, 15*(4), 143–151.

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews, 47*, 410–430.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience, 11*(4), 264–274.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*(2), 235–239.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage, 19*(4), 1835–1842.

Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences, 11*(5), 211–218.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17*(11), 565–573.

Seth, A. K. (2015). Presence, objecthood, and the phenomenology of predictive perception. *Cognitive Neuroscience* (ahead-of-print), 1–7.

Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: Predictive coding in the motor cortex. *Trends in Neurosciences, 36*(12), 706–716.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157–1162.

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655–664.