ACTIVE INTEROCEPTIVE INFERENCE AND THE EMOTIONAL BRAIN

Anil K. Seth[1] *and Karl J. Friston*[2]

[1]*Professor of Cognitive and Computational Neuroscience and Co-Director, Sackler Centre for Consciousness Science, School of Engineering and Informatics, University of Sussex, Falmer, Brighton BN1 9QJ.*
[2]*Wellcome Principal Research Fellow; Institute of Neurology, UCL, London. WC1N 3BG UK*

Correspondence: Anil Seth
Professor of Cognitive & Computational Neuroscience (Informatics)
Co-Director (Sackler Centre for Consciousness Science)
Email: A.K.Seth@sussex.ac.uk
Phone: 01273 678549

Abstract

We review a recent shift in conceptions of interoception and its relationship to hierarchical inference in the brain. The notion of interoceptive inference means that bodily states are regulated by autonomic reflexes that are enslaved by descending predictions from deep generative models of our internal and external milieu. This re-conceptualization illuminates several issues in cognitive and clinical neuroscience with implications for experiences of selfhood, and emotion. We first contextualize interoception in terms of active (Bayesian) inference in the brain; highlighting its enactivist (embodied) aspects. We then consider the key role of uncertainty or precision and how this might translate into neuromodulation. We next examine the implications for understanding the functional anatomy of the emotional brain, surveying recent observations on agranular cortex. Finally, we turn to theoretical issues; namely, the role of interoception in shaping a sense of embodied self and feelings. We will draw links between physiological homeostasis and allostasis, early cybernetic ideas of predictive control, and hierarchical generative models in predictive processing. The explanatory scope of interoceptive inference ranges from explanations for autism and depression, through to consciousness. We offer a brief survey of these exciting developments.

Keywords: emotion; self; interoception; Bayesian; predictive coding; precision; neuromodulation; depression; fatigue; autism; cybernetics; active inference.

Introduction

Recent years have seen the emergence of a framework within cognitive neuroscience that offers exactly the right set of concepts to talk about the body and mind in terms of beliefs about the body (and oneself). On this view, the brain is not an elaborate stimulus-response link but a statistical organ that actively generates explanations for the stimuli it encounters – in terms of hypotheses that are tested against sensory evidence. This perspective can be traced back to Helmholtzian formulations of unconscious inference (1). Over the last years, the underlying idea has been formalised to cover deep or hierarchical Bayesian inference – about the hidden causes of our sensations – and how these inferences induce beliefs and behaviour (2-6). 'Explanations', 'hypotheses' and 'beliefs' should in this context be understood not as consciously held mental states, but as neuronally encoded probability distributions (i.e., Bayesian beliefs) over the hidden causes of sensory signals. The biophysical encoding of these 'beliefs' is, technically, in terms of sufficient statistics like the mean or expectation of a distribution.

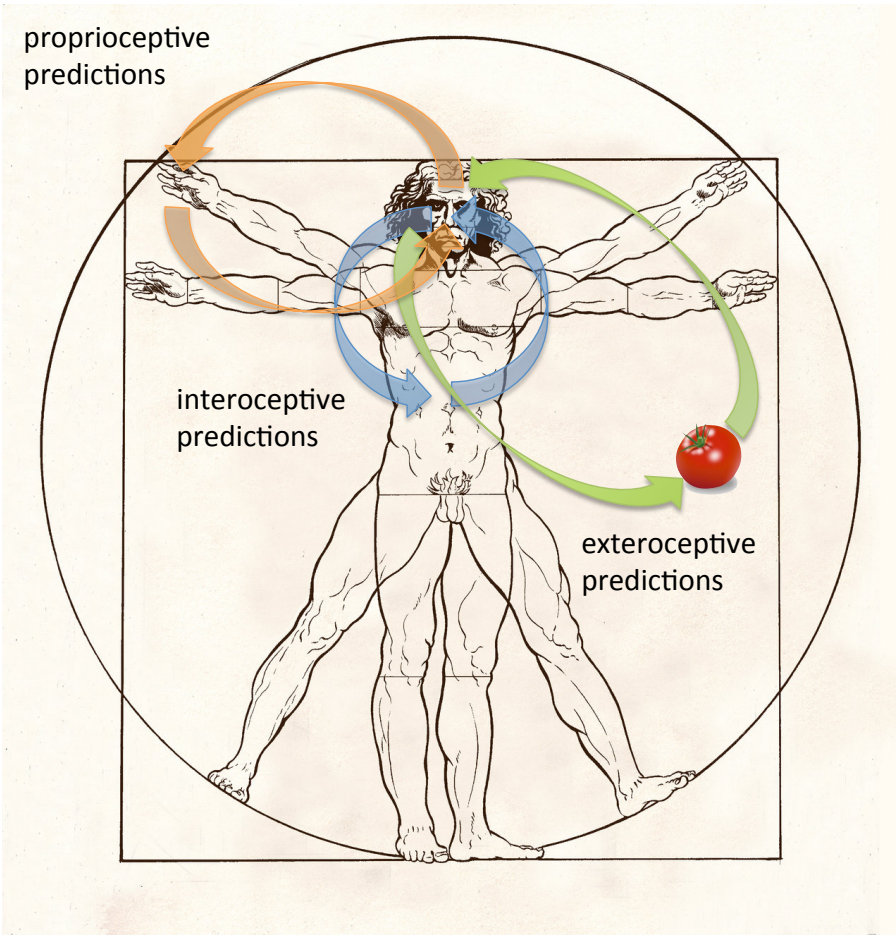

Figure 1: Inference and perception across different modalities. Green arrows represent exteroceptive predictions and prediction errors underlying perception of the external world. Orange arrows represent proprioceptive predictions (and prediction errors) generating action through active inference. Blue arrows represent interoceptive predictions (and prediction errors) underlying emotional processing and autonomic

regulation. Integrated experiences of embodied selfhood emerge from the joint hierarchical content of self-related predictions across all these dimensions, including – at hierarchically deep levels – multimodal and amodal predictions. Adapted from (7).

In the last few years, 'Bayesian brain' ideas have been applied in the context of *interoception* (see Figure 1), which refers to the perception and integration of autonomic, hormonal, visceral and immunological signals (8, 9) – or more informally as the sense of the body 'from within'. On some of these views (7, 10), emotional experience and experiences of embodied selfhood emerge from top-down inference on the (multimodal) causes of interoceptive afferents, generalizing so-called two-factor or evaluative theories of emotion and cognition (11). A first implication of these proposals is that these kinds of perceptual experience are as subject to (implicit and perhaps idiosyncratic) beliefs, as are perceptions of the external world. Beyond this, the context of interoception brings about further shifts in how to think about the relations between body, mind, and brain. One such shift is that generative models of interoceptive signals should be geared towards *control* or *regulation* of physiological variables, rather than towards accurate representation of some extra-cranial state-of-affairs (7, 12) (13). The two goals remain tightly interwoven, inasmuch as effective regulation depends on deployment of sufficiently elaborated predictive models. This shift recognizes alternative origins to 'Bayesian brain' ideas in 20th Century 'cybernetics' (14, 15) and in doing so, points to a deep connection between *life* and *mind*, in which cognitive processes are grounded in fundamental evolutionary imperatives to maintain physiological homeostasis (7, 16). Many other specific implications follow, for example in reframing the functional basis of a variety of disorders of emotion and selfhood.

Here, we survey these exciting developments. We first provide a brief introduction to the framework of prediction error minimization in the Bayesian brain, emphasizing its embodied or enactive aspects. These aspects appear prominently in the role of action in reducing prediction error (i.e., active inference) and emphasise the key role of uncertainty or precision in shaping the interplay between prior beliefs and sensory evidence. Precision-weighting of recurrent signalling in cortical hierarchies is closely associated with neuromodulation, providing important clues about developmental origins of conditions like autism; it is also associated with attention, suggesting novel accounts for symptom expression due to aberrant attention to interoceptive signals.

Turning to functional neuroanatomy, we outline the functional architecture of interoceptive inference and review recent suggestions that perceptual predictions originate preferentially in agranular cortices (9) (17) while acknowledging that direct empirical evidence for interoceptive inference is still to be uncovered. We next address some theoretical issues, relating active interoceptive inference to experiences of emotion and embodied selfhood, highlighting a control-oriented or instrumental perspective on interoceptive inference that calls on cybernetic concepts of predictive regulation, allostatic control and perceptual control theory (7, 13) (18). We conclude by exploring the implications of these ideas for a sample of clinical conditions that may reflect false interoceptive inference; either in their aetiology and/or in symptom expression. While disorders in emotional processing and interoceptive experience naturally invite explanations in terms of

abnormal interoceptive inference, we also highlight how this perspective can illuminate other conditions and symptoms including autism, fatigue, and depression.

Predictive coding in the Bayesian brain

Current formulations of Helmholtz's notion are now the most popular metaphors for neuronal processing and are usually considered under the Bayesian brain hypothesis as predictive coding (6, 19-21). Predictive coding is a process theory with a biologically plausible back story – and a considerable amount of empirical support (21, 22). See (23) for a review of canonical microcircuits and predictive coding in perception (24, 25) for an application of the same ideas to motor control, and (26) for evidence of feedforward and feedback signalling carried by distinct frequency bands.

In these schemes, neuronal representations in higher or deeper levels of neuronal hierarchies generate predictions of representations in lower levels. These descending predictions are compared with lower-level representations to form a prediction error (usually associated with the activity of superficial pyramidal cells). This mismatch or difference signal is passed back up the hierarchy, to update higher representations (usually associated with the activity of deep pyramidal cells). The recurrent exchange of signals between adjacent hierarchal levels resolves prediction error at each and every level, resulting in a hierarchically deep explanation for sensory inputs. In computational terms, the activity of neuronal populations is assumed to encode Bayesian beliefs or probability distributions over states in the world that cause sensations (e.g., my visual sensations are caused by a *face* – see Figures 2 and 3). The simplest encoding corresponds to representing the belief with the expected value (mean) of a (hidden) cause or *expectation*. These causes are referred to as *hidden* because they have to be inferred from their sensory consequences. In other words, they can never be directly observed and are forever hidden behind a sensory veil.

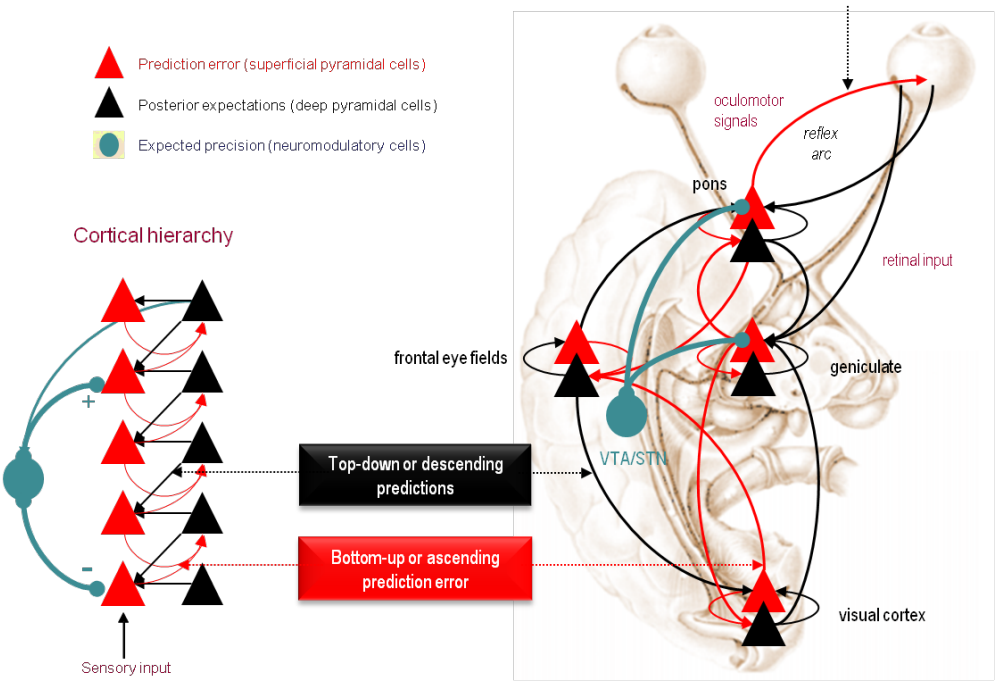

Figure 2: This figure summarizes the hierarchical neuronal message passing that underlies predictive coding. The basic idea is that neuronal activity encodes *expectations* about the causes of sensory input, where these expectations minimize prediction *error*. Prediction error is the difference between (ascending) sensory input and (descending) predictions of that input. This minimization rests upon recurrent neuronal interactions between different levels of the cortical hierarchy. Current interpretations suggest that superficial pyramidal cells (red triangles) compare the expectations (at each level) with top-down predictions from deep pyramidal cells (black triangles) of higher levels (22, 23). **Left panel**: this schematic shows a simple cortical hierarchy with ascending prediction errors and descending predictions. This graphic includes neuromodulatory gating or gain control (blue) of superficial pyramidal cells that determines their relative influence on deep pyramidal cells encoding expectations through modulation of expected precision (see below and text for details). **Right panel**: this provides a schematic example in the visual system: it shows the putative cells of origin of ascending or forward connections that convey prediction errors (red arrows) and descending or backward connections (black arrows) that construct predictions. The prediction errors are weighted by their expected precision that we have associated with the activity of neuromodulatory systems – here projections from ventral tegmental area (VTA) and substantia nigra (SN). In this example, the frontal eye fields send predictions to primary visual cortex, which it projects to the lateral geniculate body. However, the frontal eye fields also send proprioceptive predictions to pontine nuclei, which are passed to the oculomotor system to cause movement through classical reflexes. These descending predictions are also passed to the lateral geniculate body – and constitute corollary discharge. Every top-down prediction is reciprocated with a bottom-up prediction error to ensure predictions are constrained by sensory information. The resolution of *proprioceptive* prediction error is particularly important because this enables descending predictions – about the state of the body – to cause movement by dynamically resetting the equilibrium or set-point of classical reflexes. Resolving sensory prediction errors through action is known as *active inference* (see text). Adapted from (27).

In short, predictive coding represents a biologically plausible scheme for updating beliefs about the world based on sensory samples (Figure 2). In this setting, neuroanatomy and neurophysiology can the regarded as a distillation of statistical or causal structure in the environment that is disclosed by

sensory samples. The resulting anatomy of connections and their physiology furnish a generative model – generating predictions of sensations that can be compared with actual sensory samples. Empirical evidence is now emerging that shows how prior expectations shape behavioural and neuronal signatures of perception, with recent studies in vision (28-30) and audition (31) providing excellent examples. More generally, this view of perception emphasises 'the beholder's share'. See also Figure 3:

"The insight that the beholder's perception involves a top-down inference convinced [the art historian Ernst] Gombrich that there is no "innocent eye": that is, all visual perception is based on classifying concepts and interpreting visual information. One cannot perceive that which one cannot classify." (32)

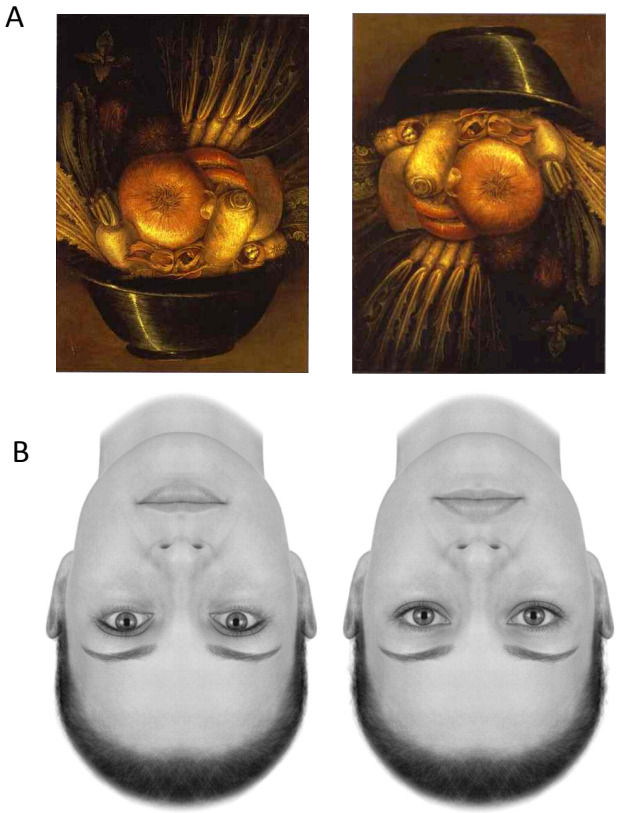

Figure 3: A. Giuseppe Arcimboldo, The Vegetable Gardener (c.1590). Oil on panel. Our percepts are constrained by what we expect to see and the hypotheses that can be called upon to explain sensory input (33). Arcimboldo, "a 16th century Milanese artist who was a favourite of the Viennese, illustrates this dramatically by using fruits and vegetables to create faces in his paintings. When viewed right side up, the paintings are readily recognisable faces." (32). Adapted from (27). B. Faces are probably one of the most important (hidden) causes of our sensations. While in Arcimboldo's image, viewing right side up is needed for the configuration of features to appear as a face, when images are already recognisably faces, viewing right side up (by rotating the page) reveals that these faces might in fact be more different than they appear (this is

the so-called "Thatcher illusion"). These examples illustrate the complex interplay between prior expectations and stimulus features that shape perceptual content [image adapted from (34)].

Embodied (active) inference and precision-weighting

There are two key ways in which prediction errors can be reduced: The first is by updating predictions to make them more like the expectations at lower levels (and sensations) currently in play. This process corresponds to perception, as implemented in predictive coding. The second way to resolve prediction errors is to change the sensory samples to make them more like predictions. This entails an active sampling of the sensorium through a redeployment of sensory surfaces; e.g., saccadic eye searches or other sensory palpitations. Placing predictive coding in an embodied or enactive framework, in which both action and perception are in the game of minimising the same prediction error is known as *active inference* (35). To fully appreciate the bilateral nature of active inference, one has to consider the embodied context in which predictions are made (and fulfilled). These predictions are not only about the world, but also about the body. In brief, perception can be understood as resolving (exteroceptive) prediction errors by selecting predictions that best explain sensations, while behaviour suppresses (proprioceptive) prediction error by changing (proprioceptive) sensations. This suppression rests on classical reflexes, whose equilibrium points are set by descending proprioceptive predictions (24). For example, an intended movement can be elicited by simply predicting the proprioceptive consequences of a particular movement trajectory, which will be fulfilled by peripheral reflexes. Note that only proprioceptive prediction errors are minimised (at the level of the spinal cord); however, with a good generative model, these movements will also fulfil visual and other exteroceptive (e.g., somatosensory) predictions. This follows because descending (multimodal) predictions emanate from a deep generative model that effectively assimilates prediction errors from all modalities – including interoception. In this context, an important and sometimes overlooked aspect of active inference is that it implies a counterfactual or conditional aspect. That is, in order for an action successfully to reduce prediction error, the brain must represent not only the hidden causes of current sensory signals but also must use these representations to predict how sensory signals would change under specific actions (36). Interestingly, it has been suggested that such counterfactual aspects of perceptual prediction may underlie basic properties perceptual experience, such as 'presence' or 'objecthood' (37).

To enable predictions about the consequences of action to be fulfilled, we have to attenuate proprioceptive prediction errors – that would otherwise deliver unequivocal evidence that we are not, in fact, acting. This attenuation rests on reducing the *precision* of proprioceptive prediction errors. Precision can be regarded as a measure of signal-to-noise or confidence. Mathematically, precision is the inverse variance or reliability of a signal. Estimating precision speaks to a fundamental aspect of inference; namely, the encoding of precision or expected uncertainty (38-40). In other words, we have to infer both the *cause* of our sensations and the *context*, in terms of the (expected or subjective) precision of sensory evidence. This represents a subtle but ubiquitous problem for the brain, where the solution rests on modulating the gain or excitability of neuronal populations reporting prediction error (21, 41, 42).

Heuristically, one can regard ascending prediction errors in cortical hierarchies as broadcasting 'newsworthy' information that cannot be explained by descending predictions. However, the brain also has to select the prediction errors it attends to. It can do this by adjusting their volume or *gain*. Those prediction errors that have been assigned high precision therefore have privileged access to high levels of the hierarchy and can therefore update high-level expectations. Empirical evidence suggests that this precision-weighting is a generic computational process throughout the brain (40) and may be instantiated through neuromodulatory mechanisms of gain control at a synaptic level (43). The ensuing neuromodulatory gain control corresponds to a (Bayes-optimal) encoding of precision in terms of the excitability of neuronal populations reporting prediction errors. This may explain why superficial pyramidal cells are equipped with so many synaptic gain control mechanisms; such as NMDA receptors and classical neuromodulatory receptors like D1 dopamine receptors (44-47). Furthermore, it places excitation-inhibition balance in a perfect position to mediate Bayesian belief updating within and among hierarchical levels (48). This contextual aspect of predictive coding has been associated with attentional gain control in sensory processing (41, 49) and has been discussed in terms of affordance in the setting of action selection (50-52). Crucially, the delicate balance of precision over hierarchical levels can have a profound effect on inference – and may underlie false beliefs in psychopathology (53).

Interoceptive inference

Key challenges for formal accounts of brain function are emotion, self-awareness and their disorders. Recently, people have started to cast emotional processing in terms of predictive coding or inference about interoceptive or bodily states (54, 55) (9, 56). The basic argument follows the explanation for action above; namely, motor reflexes are driven by proprioceptive prediction errors. Proprioceptive prediction errors compare primary afferents from stretch receptors with proprioceptive predictions that descend to alpha motor neurons in the spinal-cord and cranial nerve nuclei. This effectively replaces descending motor commands with proprioceptive predictions, which are fulfilled by peripheral reflexes (24). These predictions rest on deep hierarchical inference about states of the world, including our own body. Replacing *proprioceptive* signals with *interoceptive* signals, one can see how autonomic reflexes can transcribe descending interoceptive predictions into physiological homoeostasis (e.g., blood pressure, glycaemia, etc.). Importantly, interoceptive predictions constitute just one stream of multimodal predictions that are generated by expectations about the embodied self. On this view, interoceptive signals do not cause emotional awareness, or *vice versa*. Instead, there is a circular causality, where neuronally encoded predictions about bodily states engage autonomic reflexes through active inference (see below), while interoceptive signals inform and update these predictions. Emotion or affective content then becomes an attribute of any representation that generates interoceptive predictions – where interoception is necessarily contextualised by concurrent exteroceptive and proprioceptive cues (see Figure 1).

A useful way to think about interoceptive inference is as generalizing physiological (James-Lange) and two-factor or appraisal (e.g., (11)) approaches to emotion. These formulations regard emotional experience as arising from cognitively contextualized perception of changes in bodily state. Interoceptive inference extends these early ideas to incorporate a smooth hierarchy of (precision

weighted) predictions and prediction errors, without assuming any bright line distinction between cognitive and non-cognitive processing. By analogy with predictive coding approaches to visual perception, we propose that emotional content is determined by beliefs (i.e., posterior expectations) about the causes of interoceptive signals across multiple hierarchical levels. An important challenge in this context to is to identify which aspects of inference support specifically *conscious* emotional experience, with predictions (rather than prediction errors) being the preferred vehicle (30). It is tempting to speculate that deep expectations at higher levels of the neuronal hierarchy are candidates for – or correlates of – conscious experience; largely because their predictions are domain general and can therefore be articulated (through autonomic or motor reflexes).

Crucially, interoceptive inference augments appraisal theories with the concept of *active inference*, by which interoceptive predictions can perform physiological homeostasis by enlisting autonomic reflexes (10, 13) More specifically, descending predictions provide a homoeostatic set point against which primary (interoceptive) afferents can be compared. The resulting prediction error then drives sympathetic or parasympathetic effector systems to ensure homoeostasis or allostasis; for example, sympathetic smooth-muscle vasodilatation as a reflexive response to the predicted interoceptive consequences of 'blushing with embarrassment'. This formulation of autonomic reflexes follows exactly the active inference formulation of motor reflexes that enable the contraction of striated muscle to be prescribed or enslaved by equilibrium points set by descending projections to alpha motoneurons in the spinal-cord (57).

Active inference highlights a shift from predictive models underlying perception of hidden causes of sensory data, to their use in control or regulation of these causes (7). Importantly, both (predictive) perception and (predictive) regulation can involve action, as emphasized by distinguishing *epistemic* and *instrumental* active inference (7, 12). The basic idea is that epistemic (active) inference involves selecting actions that we expect to increase the fit between predictive models and hidden causes of sensory signals. This form of inference may characterise, for example, saccadic eye movements (36) or exploratory body movements to inform self-models (58). Instrumental active inference, by contrast, leverages predictive models to achieve *control* of sensory variables. This perspective has been applied to exteroception in the guise of 'perceptual control theory' (18) which emphasized that "control systems control what they *sense*, not what they *do*" (italics in the original). Instrumental or control-oriented inference is however particularly relevant to interoception, where maintenance of physiological variables within homeostatically viable ranges is critical to organism survival. In this context, exploratory or epistemic interoceptive 'actions' may be less evident because they may be more costly: one does not want to raise one's blood pressure to physiologically dangerous levels just to see whether it can return. The association of predictive models with control of sensory variables recalls the cybernetic view that 'every good regulator of a system must be a model of that system' (14), and the distinction between instrumental and epistemic actions also highlights the counterfactual aspects of active inference, where potential actions are associated with their likely sensory consequences (7) (37) (36).

In terms of predictive coding, the balance between homoeostatic reflexes and more goal-directed allostatic behaviour rests upon the confidence (i.e., precision), placed in deeper expectations about

how we will behave. For example, hypoglycaemia could induce low-level predictions that mobilise glucose stores (through autonomic reflexes driven by precise interoceptive prediction errors). Alternatively, if we can attenuate the precision of low-level interoception, then proprioceptive predictions can be fulfilled that preclude domain specific homoeostatic responses – and engage allostatic behaviour; i.e., preparing and consuming a meal.

As yet, direct empirical evidence for (or against) interoceptive predictions or prediction errors is still lacking. While there is ample circumstantial that fits comfortably with this framework [see (9, 10, 56, 59) for reviews], the principles of interoceptive inference rest primarily on the view that perceptual inference – whether about the world or about the body – is likely to involve a common computational architecture. Moreover, the neuroanatomical properties of brain regions involved in interoceptive processing can be informatively interpreted from this perspective, as we describe next.

Functional neuroanatomy of interoceptive inference

Translating the computational machinery of interoceptive inference into a deeper understanding of brain function requires mapping its computational elements onto neuroanatomical substrates. A number of recent proposals suggest several convergent features (9, 10, 13). The first is that so-called visceromotor areas (VMAs), such as the anterior insula cortex (AIC), anterior cingulate cortex (ACC), subgenual cortex (SGC), and perhaps also orbitofrontal cortex (OFC) are situated at the top of an interoceptive hierarchy. The second is that these areas collectively embody a generative model of interoceptive responses and issue predictions that, when unpacked at the lowest hierarchical level, serve as homeostatic set-points. These visceromotor areas are known to receive ascending projections from viscerosensory areas (e.g., posterior and mid-insula) and their descending connections engage a range of subcortical, brainstem and spinal cord targets involved in visceromotor control, such as the periaqueductal grey (PAG) and the parabrachial nucleus (PBN) (8, 60-62). Visceromotor efferents also directly innervate viscerosensory areas, potentially providing a form of efference copy or corollary discharge (i.e., descending predictions) enabling the formation of (ascending) interoceptive prediction errors. As well as known anatomical connectivity patterns, this basic architecture is supported by cytoarchitectonic observations that VMAs lack a well formed (granular) layer IV as a target for ascending prediction errors (9). Such agranular cortical regions are argued to be well suited to the issuing of predictions, in both interoceptive (9) and motor (17) domains. See Figure 4 for a schematic of the sort of functional anatomy implied by interoceptive inference (that we will appeal to later in the context of autism).

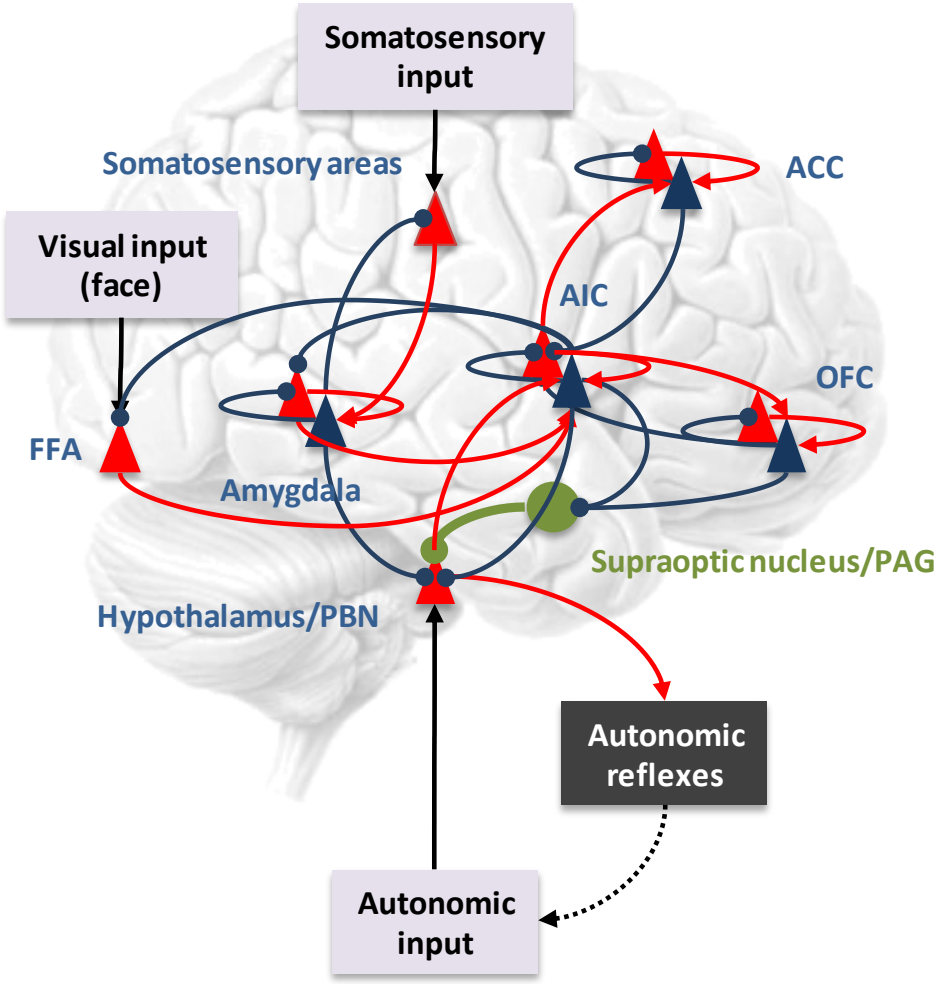

Figure 4: A (simplified) neural architecture underlying the predictive coding of visual, somatosensory and interoceptive signals. The anatomical designations, although plausible, are used to simply illustrate how predictive coding can be mapped onto neuronal systems. As in Figure 2, red triangles correspond to neuronal populations (superficial pyramidal cells) encoding prediction error, while blue triangles represent populations (deep pyramidal cells) encoding expectations. These provide descending predictions to prediction error populations in lower hierarchical levels (blue connections). The prediction error populations then reciprocate ascending prediction errors to adjust the expectations (red connections). Arrows denote excitatory connections, while circles denote inhibitory effects (mediated by inhibitory interneurons). In this example, recurrent connections mediate innate (epigenetically specified) reflexes – such as the suckling reflex – that elicit autonomic (e.g., vasovagal) reflexes in response to appropriate somatosensory input. These reflexes depend upon high-level representations predicting both the somatosensory input and interoceptive consequences. The representations are activated by somatosensory prediction errors and send interoceptive predictions to the hypothalamic area – to elicit interoceptive prediction errors that are resolved in the periphery by autonomic reflexes. Oxytocin (in green) is shown to project to the hypothalamic area, to modulate the gain or precision of interoceptive prediction error units. One hypothesis for autism rests on a failure to attenuate the precision of autonomic prediction errors, thereby precluding expectations about visual and somatosensory information (e.g., a mother's face or affiliative touch) that is not accompanied by autonomic input (see text). **FFA**; fusiform face area. **AIC**: anterior insular cortex. **ACC**: anterior cingulate cortex, **OFC**: orbitofrontal cortex. **PAG**: periaqueductal grey. **PBN**: parabrachial nucleus.

Interoceptive inference and embodied selfhood

Having described the computational architecture of interoceptive inference and its potential functional neuroanatomy, we are now in a position to explore how this framework can illuminate more theoretical issues in the nature and experience of selfhood. In everyday life, we experience our 'selfhood' as continuous and integrated. While it may be adaptive to experience being a 'self' in this way, it would be a mistake to assume on this basis that there is such a thing as unitary self-process underlying these experiences. Clinical conditions and experimental manipulations amply illustrate that experiences of selfhood unfold across many partially independent and partially overlapping levels of description; levels which can be teased apart in the laboratory or which may fall apart during psychiatric or neurological illness. A simple classification, from 'low' to 'high' levels, would range from experiences of being and having a body (10) (63) (64), through to the experience of perceiving the world from a particular point of view (a first person perspective, see (65) (66)), to experiences of intention and agency (67) (68), and at higher levels the experience of being a continuous self over time (a 'narrative' self or 'I' that depends on episodic autobiographical memory, see (69)) and finally a social self, in which my experience of being 'me' is shaped by how I perceive others' perceptions of me (70). In this putative classification, interoception plays a key role in structuring experiences of 'being and having a body' (i.e., embodied selfhood) and may also shape selfhood at other, hierarchically higher levels.

There is accumulating evidence that interoception plays a key role in shaping experiences of body ownership. Illusions of body ownership, like the rubber hand illusion and the so-called 'full body' illusion, while normally induced by false visuo-tactile congruence, can also be induced by 'cardio-visual' feedback in which a virtual body (or body part) flashes in time with a participant's heartbeat (71, 72). Recent extensions of these studies have also shown that visual feedback of respiratory patterns can have a similar effect (73), providing support for a multimodal influence of interoception on embodied selfhood. The ways in which interoceptive predictions and prediction errors shape 'higher' levels of selfhood remain exciting areas for investigation (74). As discussed next, much current evidence in these areas is found in studies of abnormal experiences of selfhood.

Selfhood and psychopathology

Very generally, the (predictive coding) process theory that we have sketched above for active inference speaks to the synaptic mechanisms that might underlie false inference in psychiatric conditions: in brief, the formal constraints implicit in predictive coding require a modulatory gain control on ascending prediction errors. A recent paper (75) exemplifies how one can understand functional (hysterical) symptoms as false inference about the causes of abnormal sensations, movements or their absence. This example offers a simple (neurophysiological) explanation of symptomatology that is otherwise rather difficult to diagnose or formulate. This theme is emerging repeatedly in psychiatry: from false inference as an account of positive symptoms (hallucinations and delusions) in schizophrenia (76), to the loss of central coherence in autism (77). Moreover, it is remarkable that the same role for precision-weighting of prediction errors emerges from different theoretical treatments of learning and inference in the brain – including predictive coding in vision

(20), free-energy accounts of perception and behaviour (4) and hierarchical Bayesian models of learning (78).

Autism and interoceptive inference

Perhaps the best example of applying concepts from interoceptive inference to understanding disorders of selfhood can be found in autism research. Recently, much of the phenomenology of autism has been cast in terms of false inference that results from a loss of prior precision, relative to sensory precision (77, 79, 80). However, in autism the consequences of increases in (or a failure to attenuate) sensory precision are also being considered in a developmental context – in which one has to accommodate the consequences for acquisition or learning of deep generative models. This is particularly interesting in relation to interoceptive inference because it touches on the acquisition of generative models that distinguish between self and other.

One line of thinking here is that a failure to contextualise interoceptive cues, elicited by interactions with the mother, precludes a proper attribution of the agency to the interoceptive consequences of prosocial interactions (18). In brief, the idea is that a failure to attenuate the precision of interoceptive prediction errors would not only render autistic infants unduly sensitive to interoceptive cues (i.e., autonomic hypersensitivity) but would have profound implications for a sense of self versus other. This follows from the inability to ignore the absence of interoceptive signals associated with nurturing (e.g., breastfeeding) during affiliative interactions with [m]others. In short, the autistic infant could never learn that the nurturing and prosocial [m]other were the same hidden cause or external object (18). See Figure 4. This has several interesting implications for attachment, theory of mind, and a lack of central coherence that characterises the disorder in later life (82). It also provides an interesting explanation for interoceptive hypersensitivity (c.f., an emotional echopraxia) in autism and failure to engage with prosocial (exteroceptive) cues (83). If this explanation is right, then it provides a clear pointer to abnormalities of (precision) gain control in cortical systems mediating interoceptive inference such as the anterior insular and cingulate cortex (84, 85).

Potential interoceptive abnormalities in autism are unlikely to reside at any single level in the interoceptive hierarchy. In a recent study, a comparison of autistic individuals with controls found that autism was associated with (i) reduced objective interoceptive sensitivity, quantified using standard heartbeat detection tasks and (ii) an increased trait interoceptive sensibility, measures using subjective questionnaires, as compared to controls (86). These results can be interpreted in terms of an increased 'interoceptive trait prediction error' (ITPE) in autism; i.e., a larger mismatch between subjective expectations about interoceptive accuracy and objective interoceptive sensitivity. Interestingly, across both autistic individuals and controls, the magnitude of ITPE correlated with self-reported anxiety, recalling the early proposal of (87) which associated anxiety with an interoceptive prediction error (though not in Bayesian framework). One complication that may nuance this view is that autism often co-occurs with alexithymia (difficulties in identifying and describing one's own emotions); a recent study found that atypical interoception was associated with alexithymia not autism, though this study did not specifically consider ITPEs (88). More

generally, the heterogeneous nature of autism may exclude single process explanations and may underlie apparently inconsistencies in the current empirical data (e.g., another recent study (89) found decreased not increased subjective body awareness in autism).

Depression and fatigue

Beyond autism, interoceptive inference is emerging as a powerful framework within which to understand depression, fatigue, and their interactions. Depression exerts a profound impact on quality of life and carries a very high socio-economic cost. Fatigue is a prominent symptom across a variety of disorders and also exacts a high toll on quality and productivity of life. While depression and fatigue encompass a wide range of cognitive, behavioural, and physiological aspects, some recent albeit speculative proposals have implicated disrupted interoception in their aetiology.

In one version of this story, peripheral endocrine and immunological changes accompanying or preceding depressive onset lead to persistently imprecise ("noisy") interoceptive afferents (9, 90). This in turn leads to lower precision-weighting of (i.e., reduced attention to) ascending interoceptive signals and correspondingly greater reliance on interoceptive priors for maintaining physiological homeostasis. Given the translation of interoceptive predictions into homeostatic set points, this process could set up a positive feedback loop in which greater reliance on prior predictions generates increasingly large and unreliable interoceptive prediction errors, which in turn increases the reliance on the now dysfunctional interoceptive predictions. At some point the ensuing dyshomeostasis will tip over into fatigue and sickness behaviour that signal the initial stages of depression (9).

In another version of the story (Stephan et al., submitted for publication), while fatigue and depression are still considered as responses to the interoceptive experience of dyshomeostasis, these now take the form of metacognitive beliefs about the brain's capacity to successfully regulate bodily states (allostatic self-efficacy). Fatigue is proposed to represent an early response to dyshomeostasis that retains adaptive value (like sickness behaviours in general), while a generalized belief of low allostatic self-efficacy following prolonged (experienced) dyshomeostasis may trigger depression, in a way that recalls cognitive theories of 'learned helplessness' (91). Both these accounts of depression are supported by the involvement of agranular visceromotor cortices in the pathophysiology of depression; e.g., (92). To further refine, distinguish, and empirically test these formulations may require advanced model-based neuroimaging analyses – of the sort being developed under the rubric of 'computational psychiatry' (93-95).

Concluding remarks

Applying the framework of active inference to interoception provides a powerful set of concepts within which to conceive the neurofunctional basis of emotion, embodied selfhood, and allostatic control. The main points can be summarized as follows. Interoceptive inference parallels other applications of active inference (or prediction error minimization) in proposing that sensory areas convey ascending prediction errors that are compared with descending predictions across a

hierarchy of perceptual processing. For interoceptive inference, predictions issue from (agranular) visceromotor areas and project to viscerosensory areas (to provide corollary feedback) as well as to brainstem and subcortical areas (to engage autonomic homeostatic reflexes). Importantly, visceromotor predictions are best interpreted as providing homeostatic set-points that enslave autonomic reflexes – and guide allostatic (behavioural and physiological) responses via interoceptive prediction errors at different hierarchical levels and timescales. This perspective emphasizes the anticipatory control-oriented nature of interoceptive inference (7), recalling the role of predictive models in cybernetic theories of regulation (14, 15) as well as their counterparts in (exteroceptive) perception; e.g., perceptual control theory (18) (96).

Mapping the computational architecture of interoceptive inference to neuroanatomical substrates – and considering the key role of precision-weighting – provide the tools to connect these ideas to (i) theories of emotion and embodied selfhood and their experimental manipulation, and (ii) a range of clinical conditions which express interoceptive symptoms and/or plausibly originate via disruptions in interoceptive inference. In terms of theoretical implications, emotional feeling states can be seen as the joint content of interoceptive predictions, while embodied selfhood rests on the multimodal and amodal predictions that distinguish self-related from non-self signals via active inference. Accumulating clinical data and experimental evidence are revealing the mechanisms by which interoceptive signalling shapes experiences of self, and also of perceptions of stimuli originating from the external environment; e.g., (97) (98). However, uncovering empirical evidence that speaks directly in favour of (or against) interoceptive inference stands as an important challenge. Key predictions of the framework are that (i) descending signals from VMAs carry predictions about the causes of interoceptive signals (and, further, that in doing so they serve as homeostatic set-points); (ii) ascending signals targeting VMAs convey interoceptive prediction errors, and (iii) emotional or affective contents depend primarily on interoceptive predictions rather than prediction errors. Future research could test these predictions using advanced laminar fMRI methods to potentially distinguish 'prediction' from 'prediction error' responses (29), or by capitalising on natural variability in physiological rhythms (e.g., heartbeat variability) to model ongoing interoceptive prediction errors that might be reflected in electrophysiological signals [see (99) for a promising approach]. Microneurography techniques – which allow direct recording of peripheral nerve traffic (100) – might also provide an innovative means of isolating interoceptive prediction and prediction error signals.

Extending active inference to include autonomic reflexes and interoceptive predictions raises many further interesting questions (27). For example, can the putative role of neuromodulators (e.g., dopamine and oxytocin) in mediating the precision of prediction errors help to explain the close relationship between arousal and anxiety? What is the relationship between exteroception and interoception during self-observation – and how does this depend upon the attenuation of the precision of respective prediction errors (101)? Do von Economo cells in infragranular cortical layers convey interoceptive predictions from the insular cortex to the amygdala and other subcortical targets? (102)? How does the control-oriented nature of interoceptive inference shape the qualitative aspects of interoceptive experience, and what in general determines the conscious status

of interoceptive predictions? Key questions about hierarchical inference and the role of interoception are also being addressed in the new field of neuropsychoanalysis (103).

The practical implications of these ideas are highlighted by their application to a variety of clinical conditions in which atypical interoceptive inference may play important roles in aetiology and/or in symptom expression. Emotional disorders like alexithymia are relatively straightforward to explain in terms of atypical interoception, while more complex and heterogeneous constructs like anxiety have been considered in terms of interoceptive prediction error for more than a decade (87, 104) (105). Recent developments have focused on depression and fatigue as emerging from the interoceptive experience of chronic dyshomeostasis, whether directly or via metacognitive beliefs in inadequate allostatic self-efficacy (Stephan et al, submitted). Autism, also highly heterogeneous, seems to have a common interoceptive foundation; possibly with a developmental origin and symptom expression characterized by discrepancies between (reduced) objective interoceptive sensitivity and (enhanced) self-appraisal of interoceptive ability. Importantly, the involvement of interoceptive inference in these and other conditions – including, for instance, depersonalisation disorder, see (106) – opens new avenues for diagnosis through physiological measures and computational psychiatry approaches, and potential clinical intervention via interoceptive training and feedback.

Altogether, considering embodied selfhood through the lens of prediction error minimization brings a new way to think about an old doctrine. Rene Descartes, besides dividing the world into *res cogitans* and *res extensa*, also achieved a certain notoriety for introducing the doctrine of the 'beast machine' (ca. 1694). He argued that while humans had minds directing their behaviour, non-human animals ('brutes') were nothing more than unthinking, unfeeling machines that breathe, digest, perceive and move 'like clockwork'. Now that we can see how human minds are deeply grounded in embodied physiology, and that similar functional principles may unite physiological regulation with perception of the external world and the guidance of actions and behaviour, an inversion of Descartes' doctrine seems plausible: that our subjective experiences of selfhood may arise *because of*, and not *in spite of*, the fact that we too are 'beast machines'.

Acknowledgements: AKS is grateful to the Dr Mortimer and Theresa Sackler Foundation, which supports the Sackler Centre for Consciousness Science. KJF is funded by the Wellcome Trust.

Conflict of interest statement: the authors declare no conflicts of interest.

References

1. Helmholtz H. Concerning the perceptions in general. Treatise on physiological optics. III. New York: Dover; 1866/1962.

2. Dayan P, Hinton GE, Neal R. The Helmholtz machine. Neural Computation. 1995;7:889-904.

3. Lee TS, Mumford D. Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am Opt Image Sc Vis. 2003;20:1434-48.

4. Friston K, Kilner J, Harrison L. A free energy principle for the brain. J Physiol Paris. 2006;100(1-3):70-87.

5. Hohwy J. The Predictive Mind. Oxford: Oxford University Press; 2013.

6. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 2013;36(3):181-204.

7. Seth AK. The cybernetic bayesian brain: from interoceptive inference to sensorimotor contingencies. In: Windt JM, Metzinger T, editors. Open MIND. Frankfurt A .M: MIND Group; 2015. p. 1-24.

8. Craig AD. Interoception: the sense of the physiological condition of the body. Curr Opin Neurobiol. 2003;13(4):500-5.

9. Barrett LF, Simmons WK. Interoceptive predictions in the brain. Nat Rev Neurosci. 2015;16(7):419-29.

10. Seth AK. Interoceptive inference, emotion, and the embodied self. Trends Cogn Sci. 2013;17(11):565-73.

11. Schachter S, Singer JE. Cognitive, social, and physiological determinants of emotional state. Psychol Rev. 1962;69:379-99.

12. Seth AK. Inference to the best prediction. In: Metzinger T, Windt JM, editors. Open MIND. Frankfurt, a. M.: GER: MIND Group; 2015. p. 1-24.

13. Pezzulo G, Rigoli F, Friston K. Active Inference, homeostatic regulation and adaptive behavioural control. Prog Neurobiol. 2015;134:17-35.

14. Conant R, Ashby WR. Every good regulator of a system must be a model of that system. International Journal of Systems Science. 1970;1(2):89-97.

15. Ashby WR. Design for a brain. London, UK: Chapman and Hall; 1952.

16. Friston KJ. Life as we know it. Journal of the Royal Society, Interface / the Royal Society. 2013;10(86):20130475.

17. Shipp S, Adams RA, Friston KJ. Reflections on agranular architecture: predictive coding in the motor cortex. Trends Neurosci. 2013;36(12):706-16.

18. Powers WT. Behavior: The control of perception. Hawthorne, NY: Aldine de Gruyter; 1973.

19. Srinivasan MV, Laughlin SB, Dubs A. Predictive coding: a fresh view of inhibition in the retina. Proc R Soc Lond B Biol Sci. 1982;216(1205):427-59.

20. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci. 1999;2(1):79-87.

21. Friston K. Hierarchical models in the brain. PLoS Comput Biol. 2008;4(11):e1000211.

22. Mumford D. On the computational architecture of the neocortex. II. Biol Cybern. 1992;66:241-51.

23. Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical microcircuits for predictive coding. Neuron. 2012;76(4):695-711.

24. Adams RA, Shipp S, Friston KJ. Predictions not commands: active inference in the motor system. Brain Struct Funct 2012;Epub ahead of print.

25. Shipp S, Adams RA, Friston KJ. Reflections on agranular architecture: predictive coding in the motor cortex. Trends Neurosci. 2013;36(12):706-16.

26. Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, et al. Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron. 2015;85(2):390-401.

27. Friston K. The fantastic organ. Brain. 2013;136(4):1328-32.

28. Kok P, de Lange FP. Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. Curr Biol. 2014;24(13):1531-5.

29. Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, et al. Contextual Feedback to Superficial Layers of V1. Curr Biol. 2015;25(20):2690-5.

30. Pinto Y, van Gaal S, de Lange FP, Lamme VA, Seth AK. Expectations accelerate entry of visual stimuli into awareness. J Vis. 2015;15(8):13.

31. Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibanez A, et al. Expectation and attention in hierarchical auditory prediction. J Neurosci. 2013;33(27):11194-205.

32. Kandel E. The Age of Insight: The Quest to Understand the Unconscious in Art, Mind, and Brain, from Vienna 1900 to the Present. New York: The Random House Publishing Group; 2012.

33. Gregory RL. Perceptual illusions and brain models. Proc R Soc Lond B. 1968;171:179-96.

34. Little AC, Jones BC, DeBruine LM. The many faces of research on face perception. Philos Trans R Soc Lond B Biol Sci. 2011;366(1571):1634-7.

35. Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. Action and behavior: a free-energy formulation. Biological Cybernetics. 2010;102(3):227-60.

36. Friston KJ, Adams RA, Perrinet L, Breakspear M. Perceptions as hypotheses: saccades as experiments. Frontiers in Psychology. 2012;3:151.

37. Seth AK. A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. Cogn Neurosci. 2014;5(2):97-118.

38. Yu AJ, Dayan P. Uncertainty, neuromodulation and attention. Neuron. 2005;46(4):681-92.

39. Brown H, Adams RA, Parees I, Edwards M, Friston K. Active inference, sensory attenuation and illusions. Cogn Process. 2013;[Epub ahead of print].

40. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HEM, et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 2013;80:519-30.

41. Feldman H, Friston KJ. Attention, uncertainty, and free-energy. Frontiers in Human Neuroscience. 2010;4:215.

42. Clark A. The many faces of precision. Front Psychol. 2013;4:270.

43. Moran RJ, Campo P, Symmonds M, but Stephan KE, Dolan RJ, Friston KJ. Free energy, precision and learning: the role of cholinergic neuromodulation. J Neurosci. 2013;33(19):8227-36.

44. Goldman-Rakic PS, Lidow MS, Smiley JF, Williams MSTaodimahpcJNTS, :. The anatomy of dopamine in monkey and human prefrontal cortex. J Neural Transm Suppl. 1992;36:163-77.

45. Braver TS, Barch DM, Cohen JD. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. Biol Psychiatry. 1999;46(3):312-28.

46. Doya K. Modulators of decision making. Nat Neurosci. 2008;11(4):410-6.

47. Lidow MS, Goldman-Rakic PS, Gallager DW, Rakic P. Distribution of dopaminergic receptors in the primate cerebral cortex: quantitative autoradiographic analysis using [3H]raclopride, [3H]spiperone and [3H]SCH23390. Neuroscience. 1991;40(3):657-71.

48. Humphries MD, Wood R, Gurney K. Dopamine-modulated dynamic cell assemblies generated by the GABAergic striatal microcircuit. Neural Netw. 2009;22(8):1174-88.

49. Jiang J, Summerfield C, Egner T. Attention sharpens the distinction between expected and unexpected percepts in the visual brain. J Neurosci. 2013;33(47):18438-47.

50. Frank MJ, Scheres A, Sherman SJ. Understanding decision-making deficits in neurological conditions: insights from models of natural action selection. Philos Trans R Soc Lond B Biol Sci. 2007;362(1485):1641-54.

51. Cisek P. Cortical mechanisms of action selection: the affordance competition hypothesis. Philos Trans R Soc Lond B Biol Sci. 2007;362(1485):1585-99.

52. Friston KJ, Shiner, T, Fitzgerald T, Galea JM, Adams R, Brown H, Dolan RJ, et al. Dopamine, affordance and active inference. PLoS Comput Biol. 2012;8(1):e1002327.

53. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Front Psychiatry. 2013;4:47.

54. Seth AK, Suzuki K, Critchley HD. An interoceptive predictive coding model of conscious presence. Front Psychol. 2011;2:395.

55. Seth AK. Interoceptive inference, emotion, and the embodied self. Trends Cogn Sci. 2013;17(11):565-73.

56. Gu X, Hof PR, Friston KJ, Fan J. Anterior insular cortex and emotional awareness. J Comp Neurol. 2013;521(15):3371-88.

57. Adams RA, Shipp S, Friston KJ. Predictions not commands: active inference in the motor system. Brain Struct Funct. 2013;218(3):611-43.

58. Bongard J, Zykov V, Lipson H. Resilient machines through continuous self-modeling. Science. 2006;314(5802):1118-21.

59. Chanes L, Barrett LF. Redefining the Role of Limbic Areas in Cortical Processing. Trends Cogn Sci. 2016;20(2):96-106.

60. Critchley HD, Harrison NA. Visceral influences on brain and behavior. Neuron. 2013;77(4):624-38.

61. Mesulam MM, Mufson EJ. Insula of the old world monkey. III: Efferent cortical output and comments on function. J Comp Neurol. 1982;212(1):38-52.

62. Mufson EJ, Mesulam MM. Insula of the old world monkey. II: Afferent cortical input and comments on the claustrum. J Comp Neurol. 1982;212(1):23-37.

63. Limanowski J, Blankenburg F. Minimal self-models and the free energy principle. Front Hum Neurosci. 2013;7:547.

64. Brugger P, Lenggenhager B. The bodily self and its disorders: neurological, psychological and social aspects. Current opinion in neurology. 2014;27(6):644-52.

65. Blanke O, Metzinger T. Full-body illusions and minimal phenomenal selfhood. Trends in cognitive sciences. 2009;13(1):7-13.

66. Ehrsson HH. The experimental induction of out-of-body experiences. Science. 2007;317(5841):1048.

67. Haggard P. Human volition: towards a neuroscience of will. Nat Rev Neurosci. 2008;9(12):934-46.

68. Friston K. Prediction, perception and agency. Int J Psychophysiol. 2012;83(2):248-52.

69. Scoville WB, Milner B. Loss of recent memory after bilateral hippocampal lesions. 1957. J Neuropsychiatry Clin Neurosci. 2000;12(1):103-13.

70. Frith CD, Frith U. Mechanisms of social cognition. Annu Rev Psychol. 2012;63:287-313.

71. Suzuki K, Garfinkel SN, Critchley HD, Seth AK. Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. Neuropsychologia. 2013;51(13):2909-17.

72. Aspell JE, Heydrich L, Marillier G, Lavanchy T, Herbelin B, Blanke O. Turning the body and self inside out: Visualized heartbeats alter bodily self-consciousness and tactile perception. Psychological Science. 2013;24(12):2445-53.

73. Adler D, Herbelin B, Similowski T, Blanke O. Breathing and sense of self: visuo-respiratory conflicts alter body self-consciousness. Respiratory physiology & neurobiology. 2014;203:68-74.

74. Tajadura-Jimenez A, Tsakiris M. Balancing the "Inner" and the "Outer" Self: Interoceptive Sensitivity Modulates Self-Other Boundaries. J Exp Psychol Gen. 2013.

75. Edwards MJ, Adams RA, Brown H, Pareés I, Friston KJ. A Bayesian account of 'hysteria'. Brain. 2012:[Epub ahead of print].

76. Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci. 2009;10(1):48-58.

77. Pellicano E, Burr D. When the world becomes 'too real': a Bayesian explanation of autistic perception. Trends Cogn Sci. 2012;16(10):504-10.

78. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A Bayesian foundation for individual learning under uncertainty. Front Hum Neurosci. 2011;5:39.

79. Skewes JC, Jegindø EM, Gebauer L. Perceptual inference and autistic traits. Autism. 2014:[Epub ahead of print].

80. Gómez C, Lizier JT, Schaum M, Wollstadt P, Grützner C, Uhlhaas P, et al. Reduced predictable information in brain signals in autism spectrum disorder. Front Neuroinform. 2014;8:9.

81. Quattrocki E, Friston K. Autism, oxytocin and interoception. Neuroscience and biobehavioral reviews. 2014;47c:410-30.

82. Happe F, Frith U. The weak coherence account: detail focused cognitive style in autism spectrum disorders. J Autism Dev Disord. 2006;36:5–25.

83. Paton B, Hohwy J, Enticott PG. The rubber hand illusion reveals proprioceptive and sensorimotor differences in autism spectrum disorders. J Autism Dev Disord. 2012;42(9):1870-83.

84. Gu X, Hof PR, Friston KJ, Fan J. Anterior insular cortex and emotional awareness. J Comp Neurol. 2013;521(15):3371-88.

85. Eilam-Stock T, Xu P, Cao M, Gu X, Van Dam NT, Anagnostou E, et al. Abnormal autonomic and associated brain activities during rest in autism spectrum disorder. Brain. 2014;137(1):153-71.

86. Garfinkel SN, Tiley C, O'Keeffe S, Harrison NA, Seth AK, Critchley HD. Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. Biological psychology. 2016;114:117-26.

87. Paulus MP, Stein MB. An insular view of anxiety. Biological psychiatry. 2006;60(4):383-7.

88. Shah P, Hall R, Catmur C, Bird G. Alexithymia, not autism, is associated with impaired interoception. Cortex. 2016;81:215-20.

89. Fiene L, Brownlow C. Investigating interoception and body awareness in adults with and without autism spectrum disorder. Autism research : official journal of the International Society for Autism Research. 2015;8(6):709-16.

90. Paulus MP, Stein MB. Interoception in anxiety and depression. Brain Struct Funct. 2010;214(5-6):451-63.

91. Seligman ME. Learned helplessness. Annual review of medicine. 1972;23:407-12.

92. Price JL, Drevets WC. Neurocircuitry of mood disorders. Neuropsychopharmacology. 2010;35(1):192-216.

93. Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016;19(3):404-13.

94. Stephan KE, Bach DR, Fletcher PC, Flint J, Frank MJ, Friston KJ, et al. Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. The lancet Psychiatry. 2016;3(1):77-83.

95. Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. The lancet Psychiatry. 2014;1(2):148-58.

96. Mansell W. Control of perception should be operationalized as a fundamental property of the nervous system. Top Cogn Sci. 2011;3(2):257-61.

97. Salomon R, Ronchi R, Donz J, Bello-Ruiz J, Herbelin B, Martet R, et al. The Insula Mediates Access to Awareness of Visual Stimuli Presented Synchronously to the Heartbeat. J Neurosci. 2016;36(18):5115-27.

98. Garfinkel SN, Minati L, Gray MA, Seth AK, Dolan RJ, Critchley HD. Fear from the heart: sensitivity to fear stimuli depends on individual heartbeats. J Neurosci. 2014;34(19):6573-82.

99. Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, et al. Uncertainty in perception and the Hierarchical Gaussian Filter. Front Hum Neurosci. 2014;8:825.

100. Macefield VG, Henderson LA. 'Real-time' imaging of cortical and subcortical sites of cardiovascular control: concurrent recordings of sympathetic nerve activity and fMRI in awake subjects. J Neurophysiol. 2016:jn 00783 2015.

101. Ainley V, Tajadura-Jiménez A, Fotopoulou A, Tsakiris M. Looking into myself: changes in interoceptive sensitivity during mirror self-observation. Psychophysiology. 2012;49(11):3936-46.

102. Critchley H, Seth A. Will studies of macaque insula reveal the neural mechanisms of self-awareness? Neuron. 2012;74(3):423-6.

103. Panksepp J, Solms M. What is neuropsychoanalysis? Clinically relevant studies of the minded brain. Trends Cogn Sci. 2012;16(1):6-8.

104. Gray MA, Harrison NA, Wiens S, Critchley HD. Modulation of emotional appraisal by false physiological feedback during fMRI. PLoS One. 2007;2(6):e546.

105. Seth AK, Suzuki K, Critchley HD. An interoceptive predictive coding model of conscious presence. Frontiers in Psychology. 2011;2:395.

106. Sedeno L, Couto B, Melloni M, Canales-Johnson A, Yoris A, Baez S, et al. How do you feel when you can't feel your body? Interoception, functional connectivity and emotional processing in depersonalization-derealization disorder. PLoS One. 2014;9(6):e98769.