Neuroscience and Biobehavioral Reviews 90 (2018) 174–183

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Interoceptive inference: From computational neuroscience to clinic

Andrew P. Owens[a,b,c,*], Micah Allen[d,e], Sasha Ondobaka[e,f], Karl J. Friston[d]

[a] *Lab of Action & Body, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, UK*
[b] *Department of Brain Repair and Rehabilitation, Institute of Neurology, University College London, UK*
[c] *Autonomic Unit, National Hospital Neurology and Neurosurgery, UCL NHS Trust, London, UK*
[d] *Wellcome Trust Centre for Neuroimaging, University College London, UK*
[e] *Institute of Cognitive Neuroscience, University College London, UK*
[f] *Sobell Department for Motor Neuroscience and Movement Disorders, University College London, UK*

ARTICLE INFO

Keywords:
Active inference
Autonomic nervous system
Free-energy principle
Homeostasis
Interoception
Interoceptive inference
Predictive coding

ABSTRACT

The central and autonomic nervous systems can be defined by their anatomical, functional and neurochemical characteristics, but neither functions in isolation. For example, fundamental components of autonomically mediated homeostatic processes are afferent interoceptive signals reporting the internal state of the body and efferent signals acting on interoceptive feedback assimilated by the brain. Recent predictive coding (interoceptive inference) models formulate interoception in terms of embodied predictive processes that support emotion and selfhood. We propose interoception may serve as a way to investigate holistic nervous system function and dysfunction in disorders of brain, body and behaviour. We appeal to predictive coding and (active) interoceptive inference, to describe the homeostatic functions of the central and autonomic nervous systems. We do so by (i) reviewing the active inference formulation of interoceptive and autonomic function, (ii) survey clinical applications of this formulation and (iii) describe how it offers an integrative approach to human physiology; particularly, interactions between the central and peripheral nervous systems in health and disease.

1. Introduction

‘Interoception’ refers to afferent sensory information arising from the sensation, perception, and awareness of afferent feedback from the viscera that underwrites homeostatic functioning (Craig, 2002). The control of interoceptive stability or homeostasis (*i.e.*, autonomic nervous system regulation) can be mapped onto a hierarchical organisation; ranging from basic physiological reflexes to global cortical networks that integrate the function of the central and autonomic nervous systems (Owens et al., 2017a). Fundamental components of these homeostatic processes are afferent interoceptive signals reporting the internal state of the body and efferent signals acting on interoceptive feedback (Barrett and Simmons, 2015; Ondobaka et al., 2015a; Quattrocki and Friston, 2014; Park et al., 2014), in the form of homeostatic reflexes that are informed by somatic states represented in the central nervous system. Co-ordinated central and peripheral nervous system function is required, even at lower tiers in the hierarchy, where structures such as the spinal cord, brainstem and hypothalamus mediate autonomic outflows and descending cortical inhibition (Calejesan et al., 2000; Benarroch, 1993). For example, the periaqueductal gray (PAG), which regulates input/output of nociceptive and visceral signals, is also innervated by descending anterior cingulate cortex (ACC) projections, which can boost or inhibit pain responsivity, selectively (Calejesan et al., 2000). Moreover, chemoreceptors in the brain stem monitor arterial carbon dioxide, oxygen and hydrogen ion levels to regulate carbon dioxide, oxygen and pH perfusion *via* sympathetic and phrenic efferents. More generally, hypothalamic, pontine and medullary sympathetic and parasympathetic nuclei interact with homeostatic representations to generate effector-organ specific autonomic responses (Saper, 2002). In the cardiovascular domain, heart rate changes are related to activity in the amygdala and dorsal anterior cingulate cortex (dACC) (Janig and Habler, 2003) and during stress, amygdala activity predicts systolic contractility (Dalton et al., 2005). The amygdala, ACC and other limbic structures supply descending inputs to the hypothalamus and brainstem for emotion-related autonomic responses (Saper, 2002).

1.1. *The functional anatomy of interoception*

As key players in the functional anatomy of interoception, the ACC and insula cortex are important for the processing of interoceptive feedback and mediating autonomic responses to interoceptive information (Medford and Critchley, 2010; Damasio and Carvalho, 2013). dACC (Critchley et al., 2003) and insula cortex (Critchley et al., 2000a;

---

[*] Corresponding author at: Lab of Action & Body, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, TW20 0EXU, UK.
E-mail address: andrew.owens@rhul.ac.uk (A.P. Owens).

https://doi.org/10.1016/j.neubiorev.2018.04.017
Received 10 May 2017; Received in revised form 11 February 2018; Accepted 16 April 2018

Critchley et al., 2000b) activity reflects engagement of sympathetic nervous system activity coupled to mental and physical behaviours. The anterior and posterior insula show increased neuronal activity during respiration, isometric exercise, cold pressor and Valsalva manoeuvres (King et al., 1999; Harper et al., 2000). Increases in blood pressure positively correlate with right dACC activity (Critchley et al., 2000a), supporting findings that sympathetic responses are lateralized to the right hemisphere (Oppenheimer et al., 1992), whereas the left insular cortex is involved in parasympathetic nervous system cardiovascular regulation, as evidenced by acute left insular stroke disrupting the correlation between heart rate and blood pressure (Oppenheimer et al., 1996).

The insula has a posterior-to-anterior gradient, with initial sensory afferent information received by the posterior insula, which is then passed to the anterior insula cortex (AIC) – especially the right – where it is integrated with cognitive-affective biases and autobiographical information. This unique integrative structure has led to a variety models relating to the function of the region, ranging from general theories of consciousness and affect to a putative role as a primary viscero-sensory region (Klein et al., 2013). Accordingly, the AIC modulates homeostatic autonomic and interoceptive function *via* connections to allostatic centres (Flynn, 1999). Reduced baroreceptor tone is associated with ACC, amygdala and AIC function, whereas initiation of baroreflexes increases activity in lateral prefrontal cortex (lPFC) and posterior insula (Kimmerly et al., 2005). The mid and posterior insula are associated with somatomotor function and representations (Damasio et al., 2000) and the AIC and mid insula cortices, ACC and somatomotor cortex are functionally associated with shifting one's attention to interoceptive signals (Critchley et al., 2004). Bilateral insula cortices are activated during oesophageal stimulation (Binkofski et al., 1998) but as stimulation increases to the point of becoming painful, the right AIC is recruited (Aziz et al., 1997), illustrating how increasing interoceptive feedback will ascend the interoceptive hierarchy from bilateral insula to right AIC, as initial reporting of somatic sensory feedback escalates to a violation of homeostasis then to nociception; engaging conscious awareness. More generally, the insula is implicated in the integration of both interoceptive and exteroceptive inputs, and has been proposed to act as a core comparator underlying the generation of a multisensory embodied self (Allen et al., 2016a; Allen and Friston, 2016), which also regulates interactions between the cognitive and affective aspects of pain (Singer et al., 2009; Wiech et al., 2010; Fardo et al., 2015).

With respect to descending neural pathways, central efferent signals can drive allostatic changes in autonomic and behavioural function. During rest (Nakamura et al., 2008) and exercise (Tatterson et al., 2000; Tucker et al., 2006), perceived changes in skin temperature and thermal discomfort typically induce behavioural modifications before the recruitment of endocrine or autonomic thermostatic mechanisms (Schlader et al., 2009; Mundel et al., 2007). Behaviour-dependent increases in blood pressure are enabled and moderated by the baroreflex (Dampney et al., 2013; Dampney et al., 2002) and baroreflex dysfunction causes loss of consciousness due to cerebral hypoperfusion. The baroreflex arc ensures cerebral perfusion by mechanoreceptors in the carotid arteries and aortic arch detecting changes in arterial pressure and constantly feeding back this interoceptive information to the nucleus of the solitary tract (NTS), which synapses with the rostral ventrolateral medulla to set efferent pressor tone. During emotional or cognitive stress, the baroreflex feedback loop is disrupted by top-down cortical influences, increasing heart rate and blood pressure during steady-state physiological demands. Specifically, the aberrant cardiovascular up-regulation in the absence of allostatic demand results from suppression of low-order baroreceptor brainstem signalling by the solitary nucleus of the medulla, hippocampus, hypothalamic nuclei and prefrontal cortex (PFC) (Skinner, 1988). In summary, although the central and autonomic nervous systems are defined by unique anatomical, functional and neurochemical characteristics, they also interact in a variety of ways to maintain homeostasis. Interoceptive signalling and control spans and integrates central and peripheral homeostatic processes, as well as influencing emotional and cognitive functions (Damasio, 1999; Gray et al., 2012; Lange and James, 1922).

In the following, we propose that interoception may provide a unique window into holistic human nervous system function and dysfunction in disorders of brain, body and behaviour. Due to the scope of this proposition, we offer a formal framework – grounded in interoceptive inference – that offers a methodological foundation for generating empirical predictions. To this end, we first formulate homeostasis in terms of interoceptive inference; *via* symbiotic interoceptive and autonomic nervous system function, before describing the clinical application of this approach. We then illustrate how this formulation can offer an overarching approach to human physiology, particularly autonomically mediated systems. Finally, we will review our initial empirical findings and their relationship to interoceptive inference.

1.2. Interoceptive predictive coding – neural correlates for conscious and unconscious processes

Discrepancies between predicted and experienced interoceptive signals have been proposed as a potential cause for anxiety (Paulus and Stein, 2006). In predictive coding terms, discrepancies between 'top-down' predictions generated by the brain and incoming sensory signals from the periphery are compared to produce a 'prediction error'. Subsequent minimisation of this prediction error corresponds to a Bayes optimal estimation of how sensory signals were caused; this can be seen easily by noting that if descending predictions match sensations exactly, the predictions must have been generated by representations of the world (*i.e.* expectations) that are, in some sense, veridical. This can be formalised in terms of Bayesian inference, where the evaluation of an expectation about the world is based on prior beliefs and the likelihood of observed data.

The application of predictive coding to perceptual inference involves minimisation of unpredicted or surprising sensory signals (prediction errors) within the cortical hierarchy by the generation of top-down predictions (Fig. 1). In this setting, the prediction errors at the sensory level play the role of a likelihood (*i.e.*, reporting how unlikely the sensations were given expectations about their causes), while prediction errors at higher levels play the role of empirical priors (*i.e.*, how unlikely expectations at one level are, given expectations of the level above). It is fairly easy to show that minimising prediction errors at each and every level of the hierarchy produces a set of expectations that constitute a Bayes optimal representation of how sensations are generated (Rao and Ballard, 1999; Friston, 2008, 2010). In brief, the minimisation of prediction errors involves reciprocal exchange of signals between hierarchical levels: prediction errors ascend the hierarchy to revise expectations, which generate descending predictions that resolve or suppress prediction errors at the level below.

In biologically plausible versions of the scheme (Friston, 2008, 2010; Shipp, 2016), prediction errors are thought to be encoded by the activity of superficial pyramidal neurons, which compare expectations with predictions descending from deep-layer pyramidal neurons in higher hierarchical levels. The prediction error is then projected (*via* intrinsic or interlaminar connections) to deep pyramidal cells encoding expectations in the higher cortical level, enabling a more accurate prediction to be reciprocated. This recurrent message passing allows prediction units to produce a more accurate prediction and effectively silence prediction error.

A prediction error's strength or influence on expectations or representations as higher levels depends on its 'precision' or reliability (Fig. 1). If a prediction error is less reliable, such as vision on a foggy day, more precision or weight will be afforded to prior expectations or beliefs about the environment. This ensures Bayes optimal perception, meaning that precision determines the influence of prediction error on subsequent hierarchical cortical evidence (*i.e.*, prediction error)