Psychological Medicine

cambridge.org/psm

What is mood? A computational perspective

James E Clark[1], Stuart Watson[1,2] and Karl J Friston[3]

[1]Newcastle University, Newcastle Upon Tyne, UK; [2]Northumberland Tyne and Wear NHS Foundation Trust, Newcastle Upon Tyne, UK and [3]University College London, London, UK

Review Article

Cite this article: Clark JE, Watson S, Friston KJ (2018). What is mood? A computational perspective. *Psychological Medicine* **48**, 2277–2284. https://doi.org/10.1017/S0033291718000430

Received: 7 July 2017
Revised: 8 January 2018
Accepted: 1 February 2018
First published online: 26 February 2018

Author for correspondence: James E Clark, E-mail: j.clark7@ncl.ac.uk

Abstract

The neurobiological understanding of mood, and by extension mood disorders, remains elusive despite decades of research implicating several neuromodulator systems. This review considers a new approach based on existing theories of functional brain organisation. The free energy principle (a.k.a. active inference), and its instantiation in the Bayesian brain, offers a complete and simple formulation of mood. It has been proposed that emotions reflect the precision of – or certainty about – the predicted sensorimotor/interoceptive consequences of action. By extending this reasoning, in a hierarchical setting, we suggest mood states act as (hyper) priors over uncertainty (i.e. emotions). Here, we consider the same computational pathology in the proprioceptive and interoceptive (behavioural and autonomic) domain in order to furnish an explanation for mood disorders. This formulation reconciles several strands of research at multiple levels of enquiry.

The current predicament

Mood disorders are heterogeneous and complex and depend upon the interplay of several neuromodulator systems and genetic and epigenetic factors (Hirschfeld, 2000; Holsboer, 2000; Nestler & Carlezon, 2006; Nutt *et al.* 2007; Dowlati *et al.* 2010; Möhler, 2012). As a result, current approaches to diagnosis and classification of mood disorders suffer several shortcomings (Nesse & Stein, 2012), and advances in understanding the underlying neurobiology have been slow. We argue that progress may be facilitated by an appreciation of the dynamic and self-organising nature of neurobiological systems (Seth, 2013; Fotopoulou, 2015; Clark *et al.* 2016).

The brain is a generative organ

Traditional hypotheses propose that the neurobiological underpinnings of a variety of disorders arise from structural or functional abnormalities in the brain consequent on a combination of environmental stress and genetic vulnerabilities (Videbech & Ravnkilde, 2004; Ota & Duman, 2013; Duman, 2014). Similarly, it has been argued that pharmacotherapy may work via its effects on neurogenesis and synaptic plasticity (Santarelli *et al.* 2003; Anacker *et al.* 2011; Castrén & Hen, 2013) and that stressors may leave an indelible scar on key neurobiological systems that disrupts their structure and/or function. In this sense, the traditional hypothesis is that brain abnormalities develop throughout the lifespan and result in the onset and maintenance of psychopathology.

However, modern accounts of brain function reject the classical notion of the brain as a passive organ in favour of theories of embodiment and self-organisation (Friston, 2010; Seth, 2013; Corlett & Fletcher, 2014). Specifically, recent theoretical accounts propose that the brain works towards an equilibrium in which its environment is rendered predictable; i.e. surprise is minimised and uncertainty resolved (Friston, 2009). Essentially, the brain embodies a generative model that encodes prior beliefs about sensory input and their causes. This model generates predictions which are tested against actual input to produce prediction errors (surprise). These prediction errors are then used by the brain to revise its model of the world. In so doing, it updates its predictions and minimises prediction error (Friston, 2010).

On this view, prior beliefs about the world, as discussed above, are represented in terms of probability distributions. The sufficient statistics for these distributions may be labelled as 'expectation' and 'precision' and, if the brain embodies such priors, it follows that they must be represented in its physical activity and anatomy. It is thought that expectations (and subsequent predictions) are encoded by synaptic activity, while precision, or uncertainty, is encoded by the extent to which this activity is attenuated or amplified (Friston & Kiebel, 2009), e.g. through careful synaptic efficacy or gain control. In current predictive coding formulations of the Bayesian brain, expectation and prediction error units are thought to occupy deep and superficial pyramidal layers of cortex, respectively (Friston & Kiebel, 2009; Bastos *et al.* 2012; Shipp *et al.* 2013); such that predictions generated, from expectations, at one level of the cortical hierarchy descend to form prediction errors in superficial layers of the level below. These prediction error units then send ascending signals to update expectations.

CAMBRIDGE UNIVERSITY PRESS

Crucially, the precision or confidence placed in prediction errors is then associated with the synaptic gain or efficacy of superficial pyramidal cells – that itself depends upon interactions with inhibitory interneurons and modulatory neurotransmission. This is a brief description of hierarchical predictive coding. In what follows, we look more closely at the back story to predictive coding; namely, free energy minimisation and allostasis.

The free energy principle

The brain, like other biological systems, seeks to maintain its physiological (and psychological) state in the face of a constantly changing internal and external environment and must therefore minimise entropy over external states (where entropy is a mathematical measure of uncertainty or expected surprise). Directly computing surprise is intractable, but by appealing to variational principles, we can calculate an upper boundary on surprise, namely free energy, which systems will (or will appear to) minimise (Friston *et al.* 2006). Given that surprise is the inverse of model evidence, if the brain is minimising a free energy bound on surprise, it is necessarily trying to maximise the sensory evidence for its model of the world and is inherently self-evidencing (Hohwy, 2016). Under some simplifying assumptions, one can equate surprise with (precision weighted) prediction error. In brief, the brain can minimise (precision weighted) prediction error in three ways. First, expectations can be updated (by changing neuronal activity) so that predictions provide a better explanation for sensory inputs (Friston *et al.* 2006). Alternatively, the brain can change the world or the way it is sampled (by engaging motor and autonomic reflexes) so that sensations fall into line with predictions. This provides a simple explanation for behaviour, which becomes the fulfilment of predicted (proprioceptive and somatosensory) motor sensations. Finally, both of these processes (perception and action) can be nuanced by optimising the precision of prediction errors. In cognitive neuroscience, this optimisation has been framed in terms of attention and attenuation. In other words, attending to a sensory stream corresponds to increasing its sensory precision through appropriate synaptic gain control. Conversely, sensory attenuation corresponds to the reduction of precision by attending away from or ignoring the consequences of one's own action.

The active minimisation of free energy is known as *active inference*, which means the brain can selectively sample from data that concur with its current expectations (Friston *et al.* 2011; Pezzulo, 2012). In the interoceptive domain, the resolution of prediction errors through autonomic reflexes provides a simple account of homoeostasis. This formulation can be extended by appealing to hierarchical generative models such that predictions at higher levels pre-empt the need for homoeostasis [e.g. in a hypoglycaemic state, by attenuating the precision of prediction errors reporting hypoglycaemia, we can suspend the reflex mobilisation of glucose and act on the world (by eating) to fulfil and maintain higher level predictions] (Pezzulo *et al.* 2015). This hierarchical minimisation of prediction errors allows allostatic control over homoeostatic reflexes (Sterling & Eyer, 1988; McEwen, 1998; Ramsay & Woods, 2014). Note that the balance between *allostasis* and *homoeostasis* depends on attenuating interoceptive prediction errors, which will be an important theme in what follows. In this way, a self-regulating embodied loop of circular causality is constructed in which the brain constructs the external environment (and internal milieu) it expects to encounter, which in turn reinforces its predictions (Seth, 2014; Barrett & Simmons, 2015).

Free energy and emotion

Recent theoretical arguments have converged on the idea that emotional states reflect changes in the uncertainty about the somatic consequences of action (Joffily & Coricelli, 2013; Wager *et al.* 2015; Seth & Friston, 2016). This uncertainty refers to the precision with which motor and physiological states can be predicted. In this setting, negative emotions contextualise events that induce expectations of unpredictability, while positive emotions refer to events that resolve uncertainty and confer a feeling of control (Barrett & Satpute, 2013; Gu *et al.* 2013). This ties emotional states to the resolution of uncertainty and, through the biophysical encoding of precision, to neuromodulation and cortical gain control (Brown & Friston, 2012).

In summary, one can associate the valence of emotional stimuli with the precision of prior beliefs about the consequences of action. In this view, positively valenced brain states are necessarily associated with increases in the precision of predictions about the (controllable) future – or, more simply, predictable consequences of motor or autonomic behaviour. Conversely, negative emotions correspond to a loss of prior precision and a sense of helplessness and uncertainty about the consequences of action.

What about mood?

Any hierarchical inference relies on hyperpriors. These furnish higher level predictions of the likely value of lower level parameters. From the above, one can see that important parameters are the precisions of prediction errors at high and low levels of the hierarchy (i.e. prior and sensory precision). These precisions reflect the confidence we place in our prior beliefs relative to sensory evidence. If emotional states in the brain reflect the precision of prior beliefs about the consequences of action, then distinct neuronal populations must also encode hyperpriors. In other words, short-term fluctuations in precision (i.e. emotional fluctuations) will themselves be constrained by hyperpriors encoding their long-term average (i.e. mood).

Here, we propose that mood corresponds to hyperpriors about emotional states, or confidence about the consequences of action. In other words, mood states reflect the prior expectation about precision that nuances (emotional) fluctuations in confidence or uncertainty. If emotion reflects interoceptive precision, and is biophysically encoded by neuromodulatory gain control, then this suggests that mood is neurobiologically encoded as the set-point of neuromodulator systems that determine synaptic gain control over principal cells reporting prediction errors at different levels of the interoceptive hierarchy. This set-point is the sensitivity of responses to prediction errors and has a profound and enduring effect on subsequent inference.

When mood goes wrong

An interesting corollary of the above account is that mood becomes a two-dimensional construct – according to the sufficient statistics (i.e. mean or expectation and precision) of hyperpriors over interoceptive precision (Fig. 1). In this sense, we might conjecture that major depression occurs when the brain is certain that it will encounter an uncertain environment, i.e. the world is inherently volatile, capricious, unpredictable and uncontrollable. There are several concomitants of this state of affairs: if, *a priori*, prior beliefs are deemed imprecise, then the attenuation of interoceptive prediction errors will be

https://doi.org/10.1017/S0033291718000430 Published online by Cambridge University Press

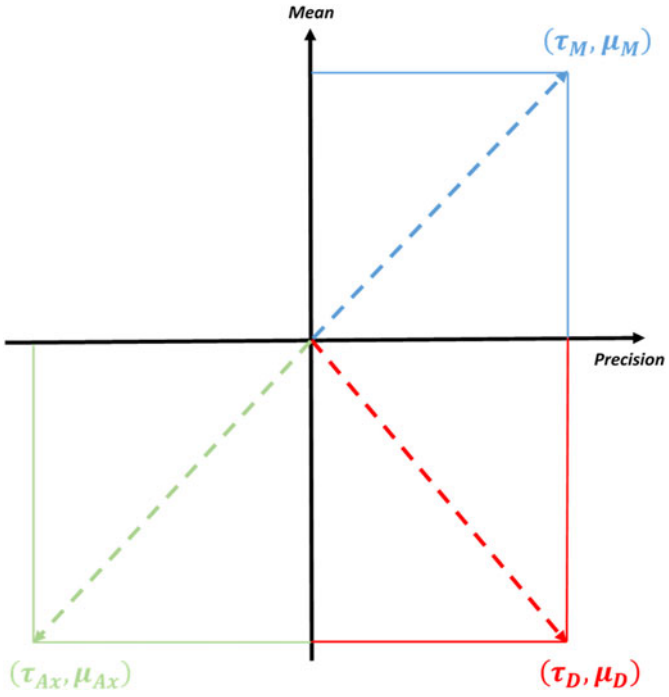

Fig. 1. The figure shows how mood can be conceptualised according to the expected precision (μ) and precision of precision (τ) in a two-dimensional space. Here, precision *per se* corresponds to the predictability of the (prosocial, affiliative and interoceptive) world – and the two dimensions correspond to hyperpriors over precision. It is proposed that pathological changes in mood occur in the extrema of this space, as highlighted. Depression occurs when an uncertain, unpredictable outcome is predicted with high precision (red lines) resulting in a chronic, self-maintaining negative emotional state that is resistant to revision. Mania (blue lines) is characterised by an equally high precision, but with the expectation of a predictable and controllable outcome – correspondingly the environment is chronically and inappropriately labelled as such. Anxiety (green lines) is an expected unpredictability but with low precision. As such, the individual engages in behaviour designed to resolve this uncertainty but which never does. D, depression; M, Mania; Ax, anxious depression.

compromised. This means that allostatic control is precluded – and is replaced by low-level homoeostatic responses (Barrett *et al.* 2016; Stephan *et al.* 2016; Peters *et al.* 2017). A corollary of this is that patients with low mood (low prior precision) should show a hypersensitivity to interoceptive cues (high sensory precision) and a failure of sensory attenuation (Badcock *et al.* 2017) – of the sort associated with stress responses (please see below).

Clearly, hyperpriors may be genetically encoded, although they may also change following chronic periods of intense stress (see below). This means the set-point of neuromodulator systems becomes configured to an aberrant tonic drive that is resistant to negative feedback loops that relay error messages (i.e. a loss of high-level precision that is reflected in persistently abnormal neuromodulation at the synaptic level). This may relate particularly to anhedonia (Chekroud, 2015). In this formalism, pleasure signals (bottom-up signals which increase interoceptive precision and confer a sense of control) – such as those generated by hedonic hotspots in hierarchically deep limbic circuitry (Berridge & Kringelbach) are attenuated – so feelings of pleasure are never initiated (Fig. 2). Such exquisite gain control – that normally allows for a precise repertoire of (stress reducing) behaviour – is denied to the depressed individual, who will fail to engage in (allostatic) actions that are likely to mitigate negative emotions. As precision enforces prediction-fulfilling action, maladaptive behaviours may therefore be conceptualised as an aberrant action-perception cycle that is self-reinforcing – or self-evidencing (Hohwy, 2016).

Mania may be associated with a comparable level of precision over expected emotional states but here, expectations are shifted towards positive emotions (of a secure, predictable, controllable and epistemically rich world). This means that the set-point of neuromodulator systems will be quantitatively distinct from depression; although they may become equally resistant to feedback. Accordingly, manic individuals lose the capacity to appreciate the unpredictable consequences of their actions and will engage in overconfident, high-risk behaviours (Mason *et al.* 2017). This has interesting parallels with theories of optimal control and reinforcement learning in which a link between action (e.g. pushing a button) and outcome (e.g. losing money) reinforce avoidance behaviour – a phenomenon which is absent or impaired in mania (see below) (Bach & Dolan, 2012).

The existence of mixed states can also be accounted for under this model. Note that mania and depression share an inappropriately precise prior over emotional states. As such, predictions encoded by such a precise prior – which do not lie at the extrema of mean values – could potentially manifest as mixed states, showing some manic and some depressive features. In Fig. 1 such states would lie in the intermediary region between the top and bottom right-hand corner.

Similarly, anxious depression can be described by a highly uncertain belief (hyperprior) about negative emotions (a loss of prior precision). This means the tonic drives of neuromodulator systems should be comparable to that in depression, but remain more responsive to peripheral feedback. One interesting consequence of this state, however, is a lack of action, given action is only possible when prior expectations are precise. In short, by framing mood as the hyperprior over emotional states, we can describe a wide range of abnormal mood states according to their different co-ordinates along a two-dimensional continuum.

The existing evidence

Neuromodulatory systems

If the account on offer is correct, we would expect an aberrant set-point for various neuromodulator systems that are specifically associated with abnormal mood states. More formally, these systems will be configured to a set-point whereby levels of stress modulators are elevated and anti-stress modulators are lowered, and receptor sensitivity is altered to ensure these levels are resistant to negative feedback from bottom-up (interoceptive) feedback from the body.

In healthy systems, mood should be affected by the valence of tightly controlled prediction errors. Recent animal work has shown that positive prediction errors (receiving more food than expected), show a strong positive correlation with dopaminergic change in the nucleus accumbens (Hart *et al.* 2014) with corresponding changes in functional brain activity in humans during a financial reward task (Rutledge *et al.* 2010). Similarly, it has been shown that signal change in the anterior insula is significantly related to the magnitude of prediction error (Bossaerts, 2010). The pharmacological manipulation of these networks was recently demonstrated where participants were given electric shocks (harms) in exchange for financial reward (gains), and offered the option of increasing the number of shocks in exchange for greater reward. It was shown that citalopram increased

https://doi.org/10.1017/S0033291718000430 Published online by Cambridge University Press

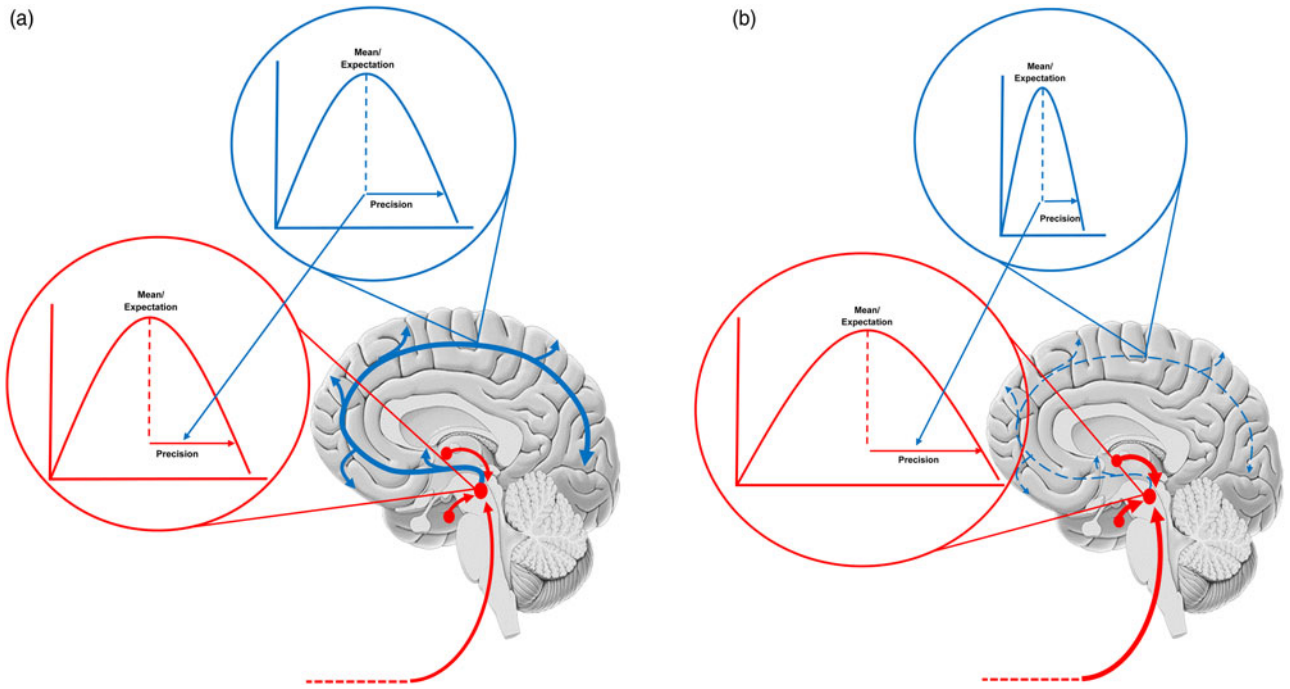

Fig. 2. The figure shows a schematic of the neuromodulatory systems with the probability distributions the embody also displayed. Ascending projections (prediction errors) are shown in red and cortical projections (predictions of precision) are in blue. The expected precision (at different levels of the cortical hierarchy) is encoded by a tonic drive – that exerts a gain control over the red (ascending) projections. Each ascending projection conveys some newsworthy (unpredicted) information. The cortical hierarchy assembles this information (i.e. prediction errors) into an updated representation of the body and world – including its predictability. Our focus in this paper is predictions of predictability (i.e. precision) that are informed by the amplitude of prediction errors from different parts of the cortical hierarchy. (a) Shows the balance in a healthy system. Mood is liable to change with environmental fluctuations due to a precision that mediates fluctuations in synaptic gain. (b) Shows how this fails in depression. A chronically stressful environment has mandated a tonically depleted serotonin drive and the estimated precision is chronically low. This precludes precise prior beliefs (and adaptive stress reducing influences from, e.g prefrontal cortex), thereby exposing cortical updating to ascending (unattenuated) autonomic drives. It is important to note this schematic is highly simplified and that similar changes may play out in other neuromodulatory systems across other mood disorders.

harm-aversion, while levodopa made individuals more likely to harm themselves than others (Crockett *et al.* 2015). This fits nicely with our notion that serotonin levels (and other neuromodulators) encode expectations about likely negative outcomes and encourage the fulfilment of these predictions through action (i.e. low levels promote behaviour with negative outcomes).

A much richer literature of neuropharmacology in mood disorders exists, and fits nicely with our theories. What is crucial is that the same systems are implicated across depression, mania and anxiety though the basal levels (expectations) and feedback sensitivity (precision) of these systems differs accordingly. Take, e.g. the hypothalamic pituitary adrenal (HPA) axis. In depression there is increased paraventricular nucleus (PVN) drive and CRH production (Raadsheer *et al.* 1994; Gao *et al.* 2013) and resistance to glucocorticoid receptor (GR) mediated negative feedback (Holsboer *et al.* 1982; Sher *et al.* 2013). We would expect that anxiety states are associated with similar basal drive but increased resistance to feedback. Indeed it has been shown that patients with post-traumatic stress disorder (PTSD) have heightened HPA axis drive originating at the amygdala (Shin *et al.* 2006) and PVN (Kasckow *et al.* 2001) and decreased inhibitory input from the hippocampus (Smith, 2005), although they show increased sensitivity to dexamethasone suppression test (Yehuda *et al.* 1993). This explains the finding of lower cortisol levels in certain testing conditions only (Meewisse *et al.* 2007). Mania is less well studied but there are reports of underactive HPA axis drive being particularly related to euphoria (Valiengo *et al.* 2012).

The effects of environmental trauma

If the brain is an organ of inference, that attempts to reduce the surprise associated with environmental outcomes, then its synaptic activity and tonic drives should be in line with its (prior and hyperprior) expectations. Active inference attempts to construct an environment (and physiology) consistent with these expectations. However, chronic unpredictability warrants a change in reliability or precision afforded to social and physiological cues. In this regard, active inference may explain the established finding that childhood trauma poses a risk for various psychopathologies (Bernet & Stein, 1999; Heim & Nemeroff, 2001; Heim *et al.* 2008). In this context, it is important to acknowledge how different types of traumatic experiences may impact differentially on predictions in the brain. Stress of any kind induces uncertainty, though the nature of this stress may determine whether the mood state will be a depressed one or an anxious one. More chronic and less well-defined adversity – which is experienced during emotional or physical neglect – is pervasive and enduring and should result in great certainty over uncertain (interoceptive) outcomes of (affiliative or prosocial) action. As such, depression would be associated with this type of trauma. In line with this, a recent study in bipolar depression demonstrated that, despite all domains of the childhood trauma being significantly more prevalent in patients than controls, only emotional neglect predicted psychopathology (Watson *et al.* 2014). Conversely, more acute and explicit traumatic experiences would change

https://doi.org/10.1017/S0033291718000430 Published online by Cambridge University Press

expectations but render them highly uncertain (i.e. induce an anxiety state). This is perhaps best exemplified by the risk posed for PTSD by acute and extreme episodes of trauma (Sullivan *et al.* 2006).

One important corollary of the theory presented above is that risk of mood disorders is intimately tied to emotional trauma only. This accords with the findings from recent studies (Watson *et al.* 2014) though not others. Instead, it is proposed that (barring brain damage) physically traumatic events facilitate the onset of mood disorders through the subjective feelings that contextualise them. One other interesting consequence of the active inference formulation is that, as mood refers to hyperpriors over interoceptive states, it is non-specific to different types of environmental fluctuations; i.e. their particular content. This means predictable environmental outcomes elsewhere can act as a buffer against the detrimental effects of trauma (Southwick & Charney, 2012). This has clear and important ramifications for preventive strategies – and may also explain why traumatic experiences are not invariably associated with the onset of psychopathology despite biologic effects (Carpenter *et al.* 2007).

The emergence of epigenetics

If environmental factors can induce conformational changes in the set-point of neurobiological systems to precipitate onset of mood disorders, then some explanatory mechanism is required. Epigenetics has emerged as a major field of enquiry in recent years, specifically in relation to methylation of the NR3C1 gene, which encodes GR (Nantharat *et al.* 2015; Smart *et al.* 2015; Palma-Gudiel *et al.* 2016). Crucially, GRs are found in the hippocampus and amygdala (Morimoto *et al.* 1996); regions that send descending predictions to the paraventricular nucleus (Herman *et al.* 2002). A loss of sensitivity to circulating cortisol levels in the amygdala and hippocampus may be the neurophysiological correlates of aberrant hyperpriors that set the neuromodulatory tone for amplification and attenuation of the PVN. Although speculative, this provides a potential framework that can be modelled in terms of active inference and is entirely consistent with theories based upon allostasis (McEwen, 2000; Radley *et al.* 2011; Braithwaite *et al.* 2015).

Our ideas also explain why epigenetic variability tends to occur at critical periods of development (Heim & Binder, 2012). In this regard it is important to remember that error signals can only be attenuated in states of high prior precision (confidence in the consequences of behaviour). Crucially, this certainty can only be inherited from a stable environment experienced over time, and so epigenetic alteration in neuromodulatory systems is more likely to occur when systems experience new or unpredictable environments and thus expect a higher degree of uncertainty. The most obvious time when this would occur is the immediate neonatal and infancy period when biological systems have almost no prior experience. The set-point of these systems, at this time, encodes an imprecise (i.e. flat or uninformative) prior that is waiting to be informed through experience. This conforms to an elegant study by Weaver *et al.* (2004) who showed a rapid increase in DNA methylation in rodents the day after birth. However, they also showed that methylation rates rapidly declined in rodents who were maternally groomed and nurtured while rates remained elevated in neglected pups (Weaver *et al.* 2004). This implicates a role of maternal care in establishing an appropriate physiological set-point and ensuring it is precise enough to resist some form of future stress challenge. It also shows that adversity experienced in *early* life is critically important in establishing vulnerability towards onset of mood disorders.

Psychological theories

Psychological theories of mood are important and remain a challenge for most biological accounts. Perhaps the best model to date is *learned helplessness* (Seligman, 1975). We have characterised depression by inappropriately high precision in the negative consequences of action for the individual's internal states. Accordingly this results in failure by the individual to engage with potentially positive outcomes of action. However, as discussed above, high precision also facilitates allostasis in order to preserve current states and so not only would depression result in failure to attend to positive stimuli but it would result in active inference to preserve the depressed state resulting in the behaviour that we associate with learned helplessness.

A more recent and promising psychological model of depression stems from the demonstration of attentional bias in patients towards negative facial expressions (Duque & Vázquez, 2015). Harmer *et al.* have thus proposed that a negative emotional bias is a core feature of depression and is the psychological target of antidepressant medication (Harmer *et al.* 2009; Harmer & Cowen, 2013) and correspondingly that modulating attentional bias can provide therapeutic benefit (Browning *et al.* 2012). Active inference requires precise coding so interoceptive information that is inconsistent with the current state can be attenuated, as such, in depression, we would expect sensory attenuation away from positive stimuli and greater attention towards negative stimuli – in line with psychological expectations that are biologically encoded.

Much experimental work has also shown an increased response to rewarding stimuli in bipolar disorder, which can be accounted for by our theory in which manic patients expect a pleasurable outcome (i.e. a reward) from their actions – even if this is highly unlikely. This has been demonstrated using self-report measures of behavioural drive (Van der Gucht *et al.* 2009) and response time analyses to financial reward cues (Singh *et al.* 2013). Accordingly patients also show a failure to learn from punishing cues (Mueller *et al.* 2010). Conversely, depressed patients show a reward hyposensitivity in line with overly precise prior prediction of a negative outcome and subsequent aversion to positive bottom-up (interoceptive) signals (Eshel & Roiser, 2010).

Future work

If the functional anatomy of mood involves alterations in tonic neuromodulatory drives, then it will affect communication between higher and lower levels of the extended interoceptive system. As such, any functional brain investigations in mood disorders must be capable of quantifying effective connectivity between relevant networks and evaluating how this connectivity is modulated by external factors. Dynamic causal modelling is well placed in this regard (Friston *et al.* 2003) and could prove a fruitful tool for further investigation – specifically in examining the task-dependent coupling between hierarchical levels of neuromodulator control (Schlosser *et al.* 2008; Lu *et al.* 2012; Radaelli *et al.* 2015; Vai *et al.* 2015; Vai *et al.* 2016).

Given the discussions above, we would expect not only differences in effective connectivity along neuromodulatory axes, but also differences in the way this connectivity is modulated

https://doi.org/10.1017/S0033291718000430 Published online by Cambridge University Press

according to external stimuli. Very few studies have been conducted along these lines but it is interesting to note a handful that have. Sladky *et al.* showed that patients with social anxiety disorder demonstrated increased activation in orbitofrontal cortex and amygdala when viewing emotional faces (Sladky *et al.* 2012), but furthermore, in controls this network shows top-down modulation – which is reversed in patients – orbitofrontal cortex drives greater activation in amygdala (Sladky *et al.* 2015). This fits nicely with our theories; where pathological anxiety is associated with an expectation of emotional negative environment (increased tonic activity in the limbic–prefrontal network) but also with a complete loss of precision and failure to dampen ascending information (loss of top-down inhibitory control). Prefrontal–amygdala connectivity in depressed patients has also been investigated in two studies, showing reduced top-down dorsolateral prefrontal modulation of amygdala response to negative images in bipolar depression (Radaelli *et al.* 2015) with a converse pattern of increased orbitomedial–amygdala connectivity when viewing positive faces (de Almeida *et al.* 2009). This is very much in line with our arguments presented above, but the imperative remains for further studies that can fully quantify neurobiologically coded expectations and precisions across different axes.

The perspective afforded by mood as a hyperprior suggests a separation of timescales in terms of responding to prediction errors. An adaptive response to a volatile environment in which the amplitude of prediction errors is, itself, on average high would suggest a mood-lowering reduction in the estimated precision of prediction errors. This is something that could, in principle, be tested experimentally using a mood induction paradigm predicated on experimentally induced prediction errors. One interesting way of achieving this may be measuring gaze duration in mood congruent and incongruent ambiguity resolving contexts, during word reading tasks.

Our theoretical account of mood may also be useful in informing molecular studies by hypothesising how alterations in interoceptive computation might play out biologically, and how this can be manipulated therapeutically. Take the example of depression. As discussed this state corresponds to the very precise expectation of a negatively valenced environment, which allows resistance to contradictory (positive) information and ensures behaviour in line with the depressed state. We are therefore presented with two possible avenues for treatment:

(1) Alter the expectation of an uncertain and uncontrollable body (or world)
(2) Alter the body (or world) to enable a revision of expected uncertainty.

These broadly concord with current treatment approaches of which the first the most widely used via pharmacological manipulations of neuromodulation. If depression – and other mood disorders – are the result of a computational pathology then no single neuromodulator system will be implicated in every patient; possibly reflecting the vast number of treatment-resistant patients and the failure to find a consistent biomarker. Higher level processing, as considered here, is the result of an array of ascending inputs and so the same computational pathology may be manifest by any pathophysiology that involves neuromodulatory systems. This speaks to figuring out ways to tailor pharmacological interventions to better match the needs of individual patients. One recent and novel study has employed machine-learning techniques to predict treatment outcome in clinical trials (Chekroud, 2015) and represents a potentially important approach; however, it may be possible to go one step further and replace the search for a biomarker with a computational signature (Nitsche *et al.* 2010; Huys *et al.* 2011; Barch *et al.* 2012; Montague *et al.* 2012; Wang & Krystal, 2014).

Conclusions

Much evidence over recent decades has converged on the idea that the brain is in the game of predicting its sensorium and working to minimise the difference between these predictions and actual sensory input. Further evidence suggests emotional states reflect the precision associated with neurobiological predictions over interoceptive states. In this paper, we have extended this formalism to a further level of the hierarchy and suggested mood acts as a hyperprior over emotional states. This notion has gained traction as an explanation for autism and schizophrenia in the exteroceptive (perceptual) domain (Lawson *et al.* 2014; Corlett, 2017; Krystal *et al.* 2017). We have explored the evidence for this theory and suggested how it might inform further research.

Acknowledgements. JEC is funded by a NIHR-RCF grant (Ref: RCF1415042). SW is an Academic Clinical Senior Lecturer with Newcastle University and an Honorary Consultant Psychiatrist with NTW Foundation Trust. KJF is funded by a Wellcome Trust Principal Research Fellowship (Ref: 088130/Z/09/Z).

Declaration of Interest. None.

References

Anacker C, Zunszain PA, Cattaneo A, Carvalho LA, Garabedian MJ, Thuret S *et al.* (2011) Antidepressants increase human hippocampal neurogenesis by activating the glucocorticoid receptor. *Molecular Psychiatry* **16**, 738–750.

Bach DR and Dolan RJ (2012) Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience* **13**, 572–586.

Badcock PB, Davey CG, Whittle S, Allen NB and Friston KJ (2017) The depressed brain: an evolutionary systems theory. *Trends in Cognitive Sciences* **21**, 182–194.

Barch DM, Carter CS, Dakin SC, Gold J, Luck SJ, Macdonald 3rd A *et al.* (2012) The clinical translation of a measure of gain control: the contrast-contrast effect task. *Schizophrenia Bulletin* **38**, 135–143.

Barrett LF, Quigley KS and Hamilton P (2016) An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20160011.

Barrett LF and Satpute AB (2013) Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology* **23**, 361–372.

Barrett LF and Simmons WK (2015) Interoceptive predictions in the brain. *Nature Reviews Neuroscience* **16**, 419–429.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P and Friston KJ (2012) Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711.

Bernet CZ and Stein MB (1999) Relationship of childhood maltreatment to the onset and course of major depression in adulthood. *Depression and Anxiety* **9**, 169–174.

Berridge KC and Kringelbach ML. (2015) Pleasure systems in the brain. *Neuron* **86**, 646–664.

Bossaerts P (2010) Risk and risk prediction error signals in anterior insula. *Brain Structure and Function* **214**, 645–653.

Braithwaite EC, Kundakovic M, Ramchandani PG, Murphy SE and Champagne FA (2015) Maternal prenatal depressive symptoms predict infant NR3C1 1F and BDNF IV DNA methylation. *Epigenetics* **10**, 408–417.

Brown H and Friston K (2012) Dynamic causal modelling of precision and synaptic gain in visual perception – an EEG study. *NeuroImage* **63**, 223–231.

Browning M, Holmes EA, Charles M, Cowen PJ and Harmer CJ (2012) Using attentional bias modification as a cognitive vaccine against depression. *Biological Psychiatry* **72**, 572–579.

https://doi.org/10.1017/S0033291718000430 Published online by Cambridge University Press

Carpenter LL, Carvalho JP, Tyrka AR, Wier LM, Mello AF, Mello MF *et al.* (2007) Decreased adrenocorticotropic hormone and cortisol responses to stress in healthy adults reporting significant childhood maltreatment. *Biological Psychiatry* **62**, 1080–1087.

Castrén E and Hen R (2013) Neuronal plasticity and antidepressant actions. *Trends in Neurosciences* **36**, 259–267.

Chekroud AM (2015) Unifying treatments for depression: an application of the free energy principle. *Frontiers in Psychology* **6**, 153–161.

Clark JE, Ng WF, Watson S and Newton JL (2016) The aetiopathogenesis of fatigue: unpredictable, complex and persistent. *British Medical Bulletin* **117**, 139–148.

Corlett PR (2017) I predict, therefore I am: perturbed predictive coding under ketamine and in schizophrenia. *Biological Psychiatry* **81**, 465–466.

Corlett PR and Fletcher PC (2014) Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *The Lancet Psychiatry* **1**, 399–402.

Crockett MJ, Siegel JZ, Kurth-Nelson Z, Ousdal OT, Story G, Frieband C *et al.* (2015) Dissociable effects of serotonin and dopamine on the valuation of harm in moral decision making. *Current Biology* **25**, 1852–1859.

de Almeida JRC, Versace A, Mechelli A, Hassel S, Quevedo K, Kupfer DJ *et al.* (2009) Abnormal amygdala-prefrontal effective connectivity to happy faces differentiates bipolar from major depression. *Biological Psychiatry* **66**, 451–459.

Dowlati Y, Herrmann N, Swardfager W, Liu H, Sham L, Reim EK *et al.* (2010) A meta-analysis of cytokines in major depression. *Biological Psychiatry* **67**, 446–457.

Duman RS (2014) Neurobiology of stress, depression, and rapid acting antidepressants: remodeling synaptic connections. *Depression and Anxiety* **31**, 291–296.

Duque A and Vázquez C (2015) Double attention bias for positive and negative emotional faces in clinical depression: evidence from an eye-tracking study. *Journal of Behavior Therapy and Experimental Psychiatry* **46**, 107–114.

Eshel N and Roiser JP (2010) Reward and punishment processing in depression. *Biological Psychiatry* **68**, 118–124.

Fotopoulou A (2015) The virtual bodily self: mentalisation of the body as revealed in anosognosia for hemiplegia. *Consciousness and Cognition* **33**, 500–510.

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

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

Friston K and Kiebel S (2009) Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **364**, 1211–1221.

Friston K, Kilner J and Harrison L (2006) A free energy principle for the brain. *Journal of Physiology – Paris* **100**, 70–87.

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

Friston KJ, Harrison L and Penny W (2003) Dynamic causal modelling. *NeuroImage* **19**, 1273–1302.

Gao S-F, Klomp A, Wu J-L, Swaab DF and Bao A-M (2013) Reduced GAD 65/67 immunoreactivity in the hypothalamic paraventricular nucleus in depression: a postmortem study. *Journal of Affective Disorders* **149**, 422–425.

Gu X, Hof PR, Friston KJ and Fan J (2013) Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology* **521**, 3371–3388.

Harmer CJ and Cowen PJ (2013) 'It's the way that you look at it' – a cognitive neuropsychological account of SSRI action in depression. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **368**, 20120407.

Harmer CJ, Goodwin GM and Cowen PJ (2009) Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. *The British Journal of Psychiatry* **195**, 102–108.

Hart AS, Rutledge RB, Glimcher PW and Phillips PEM (2014) Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *Journal of Neuroscience* **34**, 698–704.

Heim C and Binder EB (2012) Current research trends in early life stress and depression: review of human studies on sensitive periods, gene-environment interactions, and epigenetics. *Experimental Neurology* **233**, 102–111.

Heim C and Nemeroff CB (2001) The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biological Psychiatry* **49**, 1023–1039.

Heim C, Newport DJ, Mletzko T, Miller AH and Nemeroff CB (2008) The link between childhood trauma and depression: insights from HPA axis studies in humans. *Psychoneuroendocrinology* **33**, 693–710.

Herman JP, Tasker JG, Ziegler DR and Cullinan WE (2002) Local circuit regulation of paraventricular nucleus stress integration: glutamate–GABA connections. *Pharmacology Biochemistry and Behavior* **71**, 457–468.

Hirschfeld RMA (2000) History and evolution of the monoamine hypothesis of depression. *The Journal of Clinical Psychiatry* **61**, 4–6.

Hohwy J (2016) The self-evidencing brain. *Noûs* **50**, 259–285.

Holsboer F (2000) The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology* **23**, 477–501.

Holsboer F, Liebl R and Hofschuster E (1982) Repeated dexamethasone suppression test during depressive illness: normalisation of test result compared with clinical improvement. *Journal of Affective Disorders* **4**, 93–101.

Huys QJ, Moutoussis M and Williams J (2011) Are computational models of any use to psychiatry? *Neural Networks* **24**, 544–551.

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

Kasckow JW, Baker D and Geracioti TD (2001) Corticotropin-releasing hormone in depression and post-traumatic stress disorder. *Peptides* **22**, 845–851.

Krystal JH, Murray JD, Chekroud AM, Corlett PR, Yang G, Wang X-J *et al.* (2017) *Computational Psychiatry and the Challenge of Schizophrenia. Schizophrenia Bulletin* **43**, 473–475.

Lawson RP, Rees G and Friston KJ (2014) An aberrant precision account of autism. *Frontiers in Human Neuroscience* **8**, 302.

Lu Q, Li H, Luo G, Wang Y, Tang H, Han L *et al.* (2012) Impaired prefrontal-amygdala effective connectivity is responsible for the dysfunction of emotion process in major depressive disorder: a dynamic causal modeling study on MEG. *Neuroscience Letters* **523**, 125–130.

Mason L, Eldar E and Rutledge RB (2017) Mood instability and reward dysregulation – a neurocomputational model of bipolar disorder. *JAMA Psychiatry* **74**, 1275–1276.

McEwen BS (1998) Stress, adaptation, and disease: allostasis and allostatic load. *Annals of the New York Academy of Sciences* **840**, 33–44.

McEwen BS (2000) Allostasis and allostatic load: implications for neuropsychopharmacology. *Neuropsychopharmacology* **22**, 108–124.

Meewisse M-L, Reitsma JB, De Vries G-J, Gersons BPR and Olff M (2007) Cortisol and post-traumatic stress disorder in adults. *The British Journal of Psychiatry* **191**, 387–392.

Möhler H (2012) The GABA system in anxiety and depression and its therapeutic potential. *Neuropharmacology* **62**, 42–53.

Montague PR, Dolan RJ, Friston KJ and Dayan P (2012) Computational psychiatry. *Trends in Cognitive Sciences* **16**, 72–80.

Morimoto M, Morita N, Ozawa H, Yokoyama K and Kawata M (1996) Distribution of glucocorticoid receptor immunoreactivity and mRNA in the rat brain: an immunohistochemical and *in situ* hybridization study. *Neuroscience Research* **26**, 235–269.

Mueller SC, Ng P, Temple V, Hardin MG, Pine DS, Leibenluft E *et al.* (2010) Perturbed reward processing in pediatric bipolar disorder: an antisaccade study. *Journal of Psychopharmacology* **24**, 1779–1784.

Nantharat M, Wanitchanon T, Amesbutr M, Tammachote R and Praphanphoj V (2015) Glucocorticoid receptor gene (NR3C1) promoter is hypermethylated in Thai females with major depressive disorder. *Genetics and Molecular Research* **14**, 19071–19079.

Nesse RM and Stein DJ (2012) Towards a genuinely medical model for psychiatric nosology. *BMC Medicine* **10**, 1.

Nestler EJ and Carlezon WA (2006) The mesolimbic dopamine reward circuit in depression. *Biological Psychiatry* **59**, 1151–1159.

Nitsche MA, Monte-Silva K, Kuo MF and Paulus W (2010) Dopaminergic impact on cortical excitability in humans. *Reviews in the Neurosciences* **21**, 289–298.

Nutt D, Demyttenaere K, Janka Z, Aarre T, Bourin M, Canonico PL *et al.* (2007) The other face of depression, reduced positive affect: the role of catecholamines in causation and cure. *Journal of Psychopharmacology* **21**, 461–471.

https://doi.org/10.1017/S0033291718000430 Published online by Cambridge University Press

Ota KT and Duman RS (2013) Environmental and pharmacological modulations of cellular plasticity: role in the pathophysiology and treatment of depression. *Neurobiology of Disease* **57**, 28–37.

Palma-Gudiel H, Cordova-Palomera A, Feixa LM, Miquel FC and Fañanás L (2016) Epigenetic signature of glucocorticoid receptor is associated with the familial component of depression: a twin-based study. *European Psychiatry* **33**, S79.

Peters A, McEwen BS and Friston K (2017) Uncertainty and stress: why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology* **156**, 164–188.

Pezzulo G (2012) An active inference view of cognitive control. *Frontiers in Psychology* **3**, 478.

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

Raadsheer FC, Hoogendijk WJG, Stam FC, Tilders FJH and Swaab DF (1994) Increased numbers of corticotropin-releasing hormone expressing neurons in the hypothalamic paraventricular nucleus of depressed patients. *Neuroendocrinology* **60**, 436–444.

Radaelli D, Sferrazza Papa G, Vai B, Poletti S, Smeraldi E, Colombo C *et al.* (2015) Fronto-limbic disconnection in bipolar disorder. *European Psychiatry* **30**, 82–88.

Radley JJ, Kabbaj M, Jacobson L, Heydendael W, Yehuda R and Herman JP (2011) Stress risk factors and stress-related pathology: neuroplasticity, epigenetics and endophenotypes. *Stress* **14**, 481–497.

Ramsay DS and Woods SC (2014) Clarifying the roles of homeostasis and allostasis in physiological regulation. *Psychological Review* **121**, 225–247.

Rutledge RB, Dean M, Caplin A and Glimcher PW (2010) Testing the reward prediction error hypothesis with an axiomatic model. *Journal of Neuroscience* **30**, 13525–13536.

Santarelli L, Saxe M, Gross C, Surget A, Battaglia F, Dulawa S *et al.* (2003) Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science* **301**, 805–809.

Schlosser RG, Wagner G, Koch K, Dahnke R, Reichenbach JR and Sauer H (2008) Fronto-cingulate effective connectivity in major depression: a study with fMRI and dynamic causal modeling. *NeuroImage* **43**, 645–655.

Seligman MEP (1975) *Helplessness: On Depression, Development, and Death*. New York: WH Freeman/Times Books/Henry Holt & Co.

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

Seth AK (2014) The cybernetic Bayesian brain. In Open Mind. Open MIND. Frankfurt am Main: MIND Group.

Seth AK and Friston KJ (2016) Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20160007.

Sher L, Oquendo MA, Burke AK, Cooper TB and Mann JJ (2013) Combined dexamethasone suppression – corticotrophin-releasing hormone stimulation test in medication-free major depression and healthy volunteers. *Journal of Affective Disorders* **151**, 1108–1112.

Shin LM, Rauch SL and Pitman RK (2006) Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Annals of the New York Academy of Sciences* **1071**, 67–79.

Shipp S, Adams RA and Friston KJ (2013) Reflections on agranular architecture: predictive coding in the motor cortex. *Trends in Neurosciences* **36**, 706–716.

Singh MK, Chang KD, Kelley RG, Cui X, Sherdell L, Howe ME *et al.* (2013) Reward processing in adolescents with bipolar I disorder. *Journal of the American Academy of Child & Adolescent Psychiatry* **52**, 68–83.

Sladky R, Höflich A, Atanelov J, Kraus C, Baldinger P, Moser E *et al.* (2012) Increased neural habituation in the amygdala and orbitofrontal cortex in social anxiety disorder revealed by fMRI. *PLoS ONE* **7**, e50050.

Sladky R, Höflich A, Küblböck M, Kraus C, Baldinger P, Moser E *et al.* (2015) Disrupted effective connectivity between the amygdala and orbitofrontal cortex in social anxiety disorder during emotion discrimination revealed by dynamic causal modeling for fMRI. *Cerebral Cortex* **25**, 895–903.

Smart C, Strathdee G, Watson S, Murgatroyd C and McAllister-Williams RH (2015) Early life trauma, depression and the glucocorticoid receptor gene – an epigenetic perspective. *Psychological Medicine* **45**, 3393–3410.

Smith ME (2005) Bilateral hippocampal volume reduction in adults with posttraumatic stress disorder: a meta-analysis of structural MRI studies. *Hippocampus* **15**, 798–807.

Southwick SM and Charney DS (2012) The science of resilience: implications for the prevention and treatment of depression. *Science* **338**, 79–82.

Stephan KE, Manjaly ZM, Mathys CD, Weber LAE, Paliwal S, Gard T *et al.* (2016) Allostatic self-efficacy: a metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience* **10**, 550–577.

Sterling P and Eyer J (1988) Allostasis: a new paradigm to explain arousal pathology. In *Handbook of Life Stress, Cognition and Health*. (Eds K. Fisher and J. Reason.) New York: John Wiley & Sons, pp. 629–649.

Sullivan TP, Fehon DC, Andres-Hyman RC, Lipschitz DS and Grilo CM (2006) Differential relationships of childhood abuse and neglect subtypes to PTSD symptom clusters among adolescent inpatients. *Journal of Traumatic Stress* **19**, 229–239.

Vai B, Bulgarelli C, Godlewska BR, Cowen PJ, Benedetti F and Harmer CJ (2016) Fronto-limbic effective connectivity as possible predictor of antidepressant response to SSRI administration. *European Neuropsychopharmacology* **26**, 2000–2010.

Vai B, Poletti S, Radaelli D, Dallaspezia S, Bulgarelli C, Locatelli C *et al.* (2015) Successful antidepressant chronotherapeutics enhance fronto-limbic neural responses and connectivity in bipolar depression. *Psychiatry Research: Neuroimaging* **233**, 243–253.

Valiengo LL, Soeiro-de-Souza MG, Marques AH, Moreno DH, Juruena MF, Andreazza AC *et al.* (2012) Plasma cortisol in first episode drug-naïve mania: differential levels in euphoric versus irritable mood. *Journal of Affective Disorders* **138**, 149–152.

Van der Gucht E, Morriss R, Lancaster G, Kinderman P and Bentall RP (2009) Psychological processes in bipolar affective disorder: negative cognitive style and reward processing. *The British Journal of Psychiatry* **194**, 146–151.

Videbech P and Ravnkilde B (2004) Hippocampal volume and depression: a meta-analysis of MRI studies. *American Journal of Psychiatry* **161**, 1957–1966.

Wager TD, Kang J, Johnson TD, Nichols TE, Satpute AB and Barrett LF (2015) A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology* **11**, e1004066.

Wang XJ and Krystal JH (2014) Computational psychiatry. *Neuron* **84**, 638–654.

Watson S, Gallagher P, Dougall D, Porter R, Moncrieff J, Ferrier IN *et al.* (2014) Childhood trauma in bipolar disorder. *Australian and New Zealand Journal of Psychiatry* **48**, 564–570.

Weaver ICG, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR *et al.* (2004) Epigenetic programming by maternal behavior. *Nature Neuroscience* **7**, 847–854.

Yehuda R, Southwick SM, Krystal JH, Bremner D, Charney DS and Mason JW (1993) Enhanced suppression of cortisol following dexamethasone administration in posttraumatic stress disorder. *American Journal of Psychiatry* **150**, 83–83.