CellPress

Opinion

The Depressed Brain: An Evolutionary Systems Theory

Paul B. Badcock,[1,2,3,*] Christopher G. Davey,[1,3]
Sarah Whittle,[2,4] Nicholas B. Allen,[5] and Karl J. Friston[6,7]

Major depression is a debilitating condition characterised by diverse neurocognitive and behavioural deficits. Nevertheless, our species-typical capacity for depressed mood implies that it serves an adaptive function. Here we apply an interdisciplinary theory of brain function to explain depressed mood and its clinical manifestations. Combining insights from the free-energy principle (FEP) with evolutionary theorising in psychology, we argue that depression reflects an adaptive response to perceived threats of aversive social outcomes (e.g., exclusion) that minimises the likelihood of surprising interpersonal exchanges (i.e., those with unpredictable outcomes). We suggest that psychopathology typically arises from ineffectual attempts to alleviate interpersonal difficulties and/or hyper-reactive neurobiological responses to social stress (i.e., uncertainty), which often stems from early experience that social uncertainty is difficult to resolve.

Trends

The free-energy principle (FEP) is a theory of brain function asserting that action and perception operate synergistically to minimise surprise and resolve uncertainty.

Recent applications to depression have focussed on maladaptive states, concentrating either on bottom-up deficits in predictive processing or on top-down deficits in reward processing.

The relevance of evolutionary systems theory (EST) has been largely overlooked, with theorists neglecting to ask whether our capacity for depressed mood reflects an adaptive response to specific ecological challenges.

In psychology, converging lines of evidence suggest that depression stems from a need to navigate adverse social contexts.

We synthesise these paradigms (FEP and EST) by arguing that depression reflects an adaptive strategy that responds to threats of aversive interpersonal outcomes by resolving uncertainty in the social world.

An Evolutionary Systems Approach to Depression

Why do we become depressed? Why are some of us particularly prone to depression? How is this best managed? To answer these questions, we require an interdisciplinary approach that synthesises studies of the depressed brain with psychological research on its ecological, developmental, and biobehavioural correlates [1,2]. To this end, we apply an integrative **evolutionary systems theory (EST)** (see Glossary) of human brain function to explain depressed mood and its clinical manifestations. The EST in question rests on two uncontroversial assumptions. The first appeals to a consensus among cognitive scientists that the brain is a hierarchical, self-organising system sculpted by evolution [3–5]. This hierarchy ranges from lower-order, highly specialised neural subsystems responsible for sensory–motor processing to highly integrated cortical regions that develop more gradually and underlie the sophisticated, executive cognitive faculties unique to humans (Box 1). This calls for a theory of global brain function that explains how depression emerges from coordinated interactions within hierarchically integrated neuronal systems. The second assumption echoes dynamic systems approaches that situate the brain within the evolutionary dynamics of the brain–body–environment system [6–8]. According to this view, the neural mechanisms responsible for depression can be understood only by considering the broader context of human evolution, enculturation, development, embodiment, and behaviour.

We aim to exemplify this approach by offering an interdisciplinary hypothesis of the depressed brain. Following the **FEP** [4], we first discuss how depressive disorders emerge from the functioning of, and disruptions to, hierarchical neural dynamics that seek to minimise uncertainty. We then integrate this work with psychological research on the adaptive function of depression, along with the familial, developmental, and psychobiological mechanisms that often underlie it. We propose that our species-typical capacity for depressed mood can be explained as an evolved biobehavioural strategy that responds adaptively to adverse

[1]Centre for Youth Mental Health, The University of Melbourne, Melbourne, VIC 3052, Australia
[2]Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC 3010, Australia
[3]Orygen, The National Centre of Excellence in Youth Mental Health, Melbourne, VIC 3052, Australia
[4]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Melbourne, VIC 3053, Australia

 http://dx.doi.org/10.1016/j.tics.2017.01.005

interpersonal conditions by minimising the likelihood of unpredictable social interactions. We discuss how our model builds on theories of clinical depression in the **active inference** literature, before turning to the hierarchical neural mechanics that underlie depressed mood and depressive disorders.

[5]Department of Psychology, University of Oregon, Eugene, OR 97403, USA
[6]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK
[7]Senior author

*Correspondence:
pbadcock@unimelb.edu.au
(P.B. Badcock).

Applying the FEP to Depression

The FEP is a global theory of neural structure and function, which suggests that the brain can be seen as a 'prediction machine' that attempts to maximise the evidence for a creature's model of the world by minimising an upper limit on **surprise** (i.e., free-energy; Box 2). In line with **predictive coding**, the FEP describes the brain as a hierarchical **generative model** – a hierarchy of hypotheses about the world that enables a reduction of surprise by minimising discrepancies between incoming sensory inputs and top-down predictions [9]. Conditional expectations are thought to be encoded by deep pyramidal cells (i.e., representation units) at each level of the cortical hierarchy that convey predictions downward to suppress errors at the level below, while prediction errors are encoded by superficial pyramidal cells (i.e., error units) that convey errors forward to revise expectations at the level above [10]. This allows us to minimise surprise by updating our internal models (i.e., perception). Alternatively, we can selectively sample sensory data to ensure that our predictions are self-fulfilling by changing how we act on the world to confirm our expectations (i.e., active inference [11]). Thus, perception and action operate synergistically to minimise prediction errors and optimise our internal representations of the environment. A key corollary of this model is that our predictions are optimised by evolution, development and learning. Emphasis is placed on **adaptive priors** – inherited expectations about the way our world unfolds that have been shaped by natural selection to guide action–perception cycles toward adaptive (i.e., unsurprising) states [4,12].

To date, applications of the FEP to depressive disorders have chiefly concentrated on two processes, stemming from different levels of the cortical hierarchy. The first relates to limbic deficits in minimising prediction error. In this model, depressive disorders arise from aberrant **interoceptive** predictions originating from abnormalities within the (limbic) agranular visceromotor cortex, which is central to emotional processing, energy regulation, and allostatic responses to stress [13,14]. These abnormalities can arise from past exposure to sustained distress and generate false (interoceptive) predictions about the body's imminent autonomic, metabolic, and immunological needs that activate physiological stress responses, leading to

Box 1. The Hierarchical Structure of the Brain

In psychology, it has long been recognised that the brain entails a hierarchical structure ranging from highly specialised sensorimotor systems at its lowest levels to developmentally flexible, highly integrated systems responsible for higher-order executive functions [3,72]. A hierarchical neural architecture is also emphasised by predictive coding approaches in neuroscience, which explore how the brain minimises prediction error via recurrent message passing between cortical levels [9,73,74]. More recently, imaging studies in network neuroscience have provided direct evidence that the brain exhibits a multiscale hierarchical organisation, with a given node (e.g., network, module, submodule) itself comprising a network of smaller interacting nodes at a lower level [73,75,76] (Figure I).

Comparative work suggests that a hierarchical architecture is a hallmark of the mammalian brain, progressing from highly segregated sensorimotor hierarchies common to all mammals to the cortical association areas that confer the adaptive advantage of heightened cognitive control among primates [77,78]. This structure is thought to exemplify the complementary relationship between evolution and development: selection has canalised early sensorimotor regions that serve as neurodevelopmental anchors, allowing the progressive, activity-dependent self-organisation of the widely distributed association networks that lie furthest from sensory patterning centres [77,79]. This neuroplasticity enhances adaptability by producing higher-order, 'domain-general' faculties that are able to respond flexibly to rapidly changing environments [6,79].

It is now broadly agreed that a hierarchical neural structure is favoured by selection. It enhances evolvability because deleterious changes to a single component of the system are unlikely to affect the system itself and it allows adaptive