ANZJP

Australian & New Zealand Journal of Psychiatry
2024, Vol. 58(1) 5–9
DOI: 10.1177/00048674231219178

Sage

Understanding and explaining depression: From Karl Jaspers to Karl Friston

Christopher G Davey

I have had two main interests through my career: working as a psychiatrist to help patients with severe depression and exploring the neuroscience of depression with the aim of learning more about its aetiology and mechanisms of treatment response. My clinical work has helped to frame the neuroscience questions, but in truth, the research findings have done little to inform my understanding of patients. The gap between these two components of my work as a clinician-scientist has made me reflect on why such a gap exists (and why it seems so wide), and to consider how it might be bridged.

The need to learn more about depression and how to treat it is pressing. Depression causes more disability than any other mental disorder and is one of the main causes of disability across the community. Despite this – and perhaps because of it – it remains a contested diagnosis. Operationalised as 'major depressive disorder' (MDD), it is contested in terms of the breadth of its diagnostic scope, its causes and its treatments.

Many of the arguments can be characterised as an opposition between two perspectives. On the one side, depression is argued to be an understandable response to the stresses and strains of contemporary life. It can be understood as a psychological reaction to social events, and not something in need of assessment, diagnosis and treatment (i.e., it is not something that needs to be subjected to the medical model). Depression can be alleviated by attending to the social factors that have led to it, and if treatment is to be suggested, it should focus on psychological and lifestyle factors.

From another perspective, depression is explained as arising from aberrant brain processes, with many of its symptoms – fatigue, insomnia, lack of appetite, reduced sexual interest – suggesting that it has a clear physical basis. When severe, depression can have serious physical consequences, including life-threatening dehydration and starvation. Treatments should rightly include medications that have effects on mood, and other biologically focused treatments might also be considered.

'Major depressive disorder'

Depression has been a feature of human experience since the beginnings of recorded history, and we can assume before that. Our current conceptualisation of the disorder – MDD – emerged from the neo-Kraepelinian movement. The outline of the diagnosis was first formulated by a group of research psychiatrists at Washington University in St. Louis in the 1970s, and the diagnosis was given authority with the publication of the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.; DSM-III) in 1980. The criteria for MDD aimed to provide an atheoretical approach to diagnosis – one that contrasted with the psychoanalytic approach that had held sway in the mid-20th century. It was neo-Kraepelinian in that it returned to the observational perspective favoured by Emil Kraepelin, listing the symptoms associated with the disorder without resting them on any aetiological assumptions.

The diagnosis of MDD was criticised from the beginning. The early focus was on the way the diagnosis combined different forms of depression, eliding the old diagnosis of melancholia. This, Shorter (2007) said, was 'a nosological catastrophe from which . . . the field has not recovered'. There has also been criticism directed at the broadness of the diagnosis, and that it captures too much of normal variation in mood, pathologising understandable responses to the travails of life.

Many people agree that the diagnostic criteria for depression and related disorders are flawed, and propose different ideas as to how the diagnoses might be improved. The question poses itself, however: can any categorisation of mental illness properly account for the variety of presentations of mental distress? One of the early critics of the Kraepelinian approach was Karl Jaspers, a German psychiatrist who trained at Kraepelin's University of Heidelberg shortly after his retirement.

Department of Psychiatry, The University of Melbourne, Parkville, VIC, Australia

Corresponding author:
Christopher G Davey, Department of Psychiatry, The University of Melbourne, 161 Barry Street, Carlton, VIC 3053, Australia.
Email: c.davey@unimelb.edu.au

Correction (May 2024): This article has been updated with minor grammatical or style corrections since its original publication.

Karl Jaspers

Jaspers published the first edition of the book for which he is best known, *General Psychopathology*, in 1913 at the age of 30: following which he left psychiatry to pursue a career as a philosopher. He wrote it when he was a psychiatry trainee at the University of Heidelberg, in a psychiatry department that was still in thrall to Kraepelin.

Jaspers was, however, sceptical of Kraepelin's nosological focus, and thought that psychiatry first needed to get a better grasp on its conceptual foundations. He developed a descriptive psychopathology that focused on the patients' reported experiences. He also developed a framework for how psychiatrists come to know the nature of a patient's mental health difficulties. His approach rested on the notion that psychiatry works at the interface between the sciences and humanities. While Jaspers believed that biological processes were important for the generation of mental illness, he believed that people were not divisible into their constituent parts, which scientific explanations demand. People are always whole and complex, he argued, and never completely captured by any single method of knowledge.

Jaspers introduced the distinction between *understanding* and *explaining* in how psychiatrists come to know their patients. On the one hand, the psychiatrist understands the patient's difficulties through their intuitive sense of the connections between the patient's psychological experiences and their psychosocial circumstances. The psychiatrist does this by empathising with the patient's predicament, aided by their experience of having seen many patients who have had similar experiences. They can understand, for example, why a patient's mood is low after having endured a relationship break-up, and why they felt so distressed on seeing their ex-partner start a relationship with someone else. Jaspers termed this *verstehen*, translated into English as 'understanding'.

Psychiatrists also bring scientific knowledge to their interaction with their patients. Findings from research help to explain their symptoms: Jaspers referred to this knowledge as *erklären*, or in English, 'explanation'. The psychiatrist might gain insight into their patients experiences through their knowledge of research that shows that depression arises because of aberrant functioning of the serotonergic system, or because of overactivity of the amygdala and medial prefrontal cortex, or because the patient has a neurotic personality style. These research findings have been generated by studying groups of patients, and to do this patients must be classified according to their meeting diagnostic criteria.

Jaspers was working at a time when biological theories were dominant – Kraepelin believed that mental illnesses arose from biological processes, and his nosology was, in part, an effort to provide the basis for discovering them. While Jaspers also believed that biological processes were important, he thought that such explanations fall short of capturing the complexity of the experiences of our individual patients. The two ways we have of knowing our patients – understanding and explaining – approach the patient from different perspectives, never quite reaching each other. 'The situation', Jaspers (1959) said, 'is analogous with the exploration of an unknown continent from opposite directions, where the explorers never meet because of the impenetrable country that intervenes'.

That is not to say that the gap between the two cannot be narrowed by improving our explanations. While understanding is argued to apply intuitively and to the whole patient, we bring many aspects of our knowledge to this understanding, including background information about culture, history, socioeconomic processes, and so on (Ghaemi, 2013). Our knowledge of biological processes can also contribute to our understanding, and with more coherent frameworks, improve our abilities to understand our patients.

Nosology

Jaspers believed that nosological schemes have some function – allowing us to collect statistical data, for example – but argued that any scheme that attempted to place the complexity of individual patients into neat diagnostic boxes would ultimately fail. 'Classification is always contradictory in theory and never quite squares with the facts' (Jaspers, 1959).

The benefits that can flow from the classification of natural phenomena were demonstrated by Linnaeus's classification of animals and plants. The success of his taxonomy was due to a fact about nature that he was unaware of at the time he developed it. It is descent by natural selection that gives shape to the descending flow of families, genera and species. Kraepelin's nosology was compared by his contemporaries to Linnaeus's taxonomy; although as Jaspers (1959) noted, '[mental illnesses] are not like plants which we can classify in a herbarium. Rather it is just what is a "plant" – an illness – that is most uncertain'.

If Linnaeus's classification of the plant and animal kingdoms are celebrated, his taxonomy for his third kingdom, minerals, has not aged so well. His classification of minerals into rocks, ores and deposits was developed similarly to his animal and plant taxonomy. Not being subject to natural selection, however, his mineral taxonomy now makes little sense. Instead, we make sense of the geological world via a very different system of classification: by way of Mendeleev's periodic table. It is, Hacking (2013) suggests, by means that are at present unknown to us, and akin to the relationship between the periodic table and Linnaeus's taxonomy, that we will one day come to understand mental illnesses.

A network approach

Jaspers understood symptoms of depression and anxiety as being present in different combinations in each individual, as their distinct personalities interacted with their particular social circumstances. This is a perspective that has been supported by network theories of mental disorders (Borsboom, 2017). This approach considers symptoms as ends in themselves, and not as manifestations of underlying disorders. The relationships between symptoms are observed in people over time, and by examining their temporal courses inferences can be made on the causal effect that one symptom has on another. The presence of insomnia, for example, can be observed to cause the later onset of fatigue.

The network approach upends our usual approach to mental disorders, where symptoms are believed to be the observable manifestations of the disorders that we describe in our nosological systems. Network theories suggest that there are no latent disorders underlying the symptoms. While symptoms cluster together in patterns that are not random, they do not cluster in a way that supports our diagnostic criteria. Mood-related symptoms cluster with anxiety-related symptoms, and their categorisation into, for example, MDD and social anxiety disorder, is artifactual, and not a reflection of natural categories.

The gap between the symptomatic presentations of patients and our nosological systems provides one reason that Jasperian explanation falls short of understanding. Our diagnostic categories, which are necessary for research, do not match the complex presentations of our patients and have the effect of narrowing our perspectives of on patients' difficulties (Ghaemi, 2013). If there are no latent disorders underlying the symptomatic presentations of our patients – if the symptoms are all that there is – then how do we explain what is going on for them?

The predictive brain

A recent theory of brain function, when applied to interoception, provides a framework that has the capacity to narrow the understanding–explanation gap proposed by Jaspers. It is a theory that rests on theoretical neuroscience and the free energy principal developed by Karl Friston. Like Jaspers, Friston is a psychiatrist who has forged his reputation outside of the discipline: in his case in computational neuroscience, initially as applied to brain imaging data, and more recently in the development of mathematical models of living systems.

The free energy principle posits that the brain functions to minimise free energy so as to maintain an organism within its viable physiological bounds (Friston, 2010). Free energy is a quantity borrowed from machine learning and statistical thermodynamics, and for our purposes can be said to approximate the degree of informational uncertainty in a system. Minimising free energy opposes the tendency of living systems to decay, or to increase their entropy as per the second law of thermodynamics. The predictive processing account of brain function develops this principle to explain how such minimisation of free energy occurs.

Predictive processing suggests that the brain's primary mode of operation is to generate predictions about the sensory shape of the world as we act upon it. The brain predicts the sensory signals it will receive, which it compares to the actual signals, creating a 'prediction error' that is the difference between them. The brain's models should be as close a match as possible to the incoming sensory signals (its models should accurately represent the world), and to do this, the brain minimises the prediction errors. This has the effect of minimising the amount of variational free energy, thus preserving system integrity.

One way for the brain to minimise prediction errors is to update its models via perceptual inference: what looked like a stick is now seen to be moving and the model is updated to account for it being a snake. A second way of minimising prediction errors is by action, where we act upon the world to fulfil our predictions. Active inference complements perceptual inference by proposing that sensory predictions include those that pertain to proprioception and interoception, and that prediction errors in these domains can be minimised by actions that change the sensory data. Perceptual inference and active inference operate in tandem to maximise model evidence (the term *active inference* is often used to describe the whole framework).

Motor action proceeds, according to active inference, by the generation of predictions about proprioceptive sensations, with action then fulfilling those predictions. Interoception, which refers to the processes by which we sense, integrate and regulate signals from within ourselves, is hypothesised to proceed analogously. The brain makes predictions about the effects of smooth muscle contraction and neuroendocrine activity on interoceptive variables – with activity adjusted to minimise prediction errors. By this means the brain controls basic physiological parameters, such as body temperature and osmolality, and also more complex processes, such as arousal, sleep, appetite and emotions.

The predictive processing framework is hierarchical, with predictions being encoded in broad terms at higher levels and gaining greater specificity as they descend, and prediction errors that are encoded at low levels ascending in the reverse direction. Most interoceptive prediction errors are resolved without conscious awareness: generative models make firm predictions about parameters such as body temperature and plasma osmolality that need to be maintained within narrow bands, and their corresponding prediction errors are attended to at low levels of the neural hierarchy via reflex arcs. They only

arise to conscious awareness as affective experiences when the prediction errors cannot be resolved at lower levels and require behaviours to minimise them: a feeling of thirst that leads to the drinking of water, or a feeling of anxiety that creates a preparedness to flee.

Making sense of how prediction errors are processed requires introducing another component of predictive processing: the concept of precision. Precision refers to how reliable a prediction error is believed to be (it is the inverse of its variance). Our weighting of prediction errors – the degree to which we allow prediction errors to update the generative models – is influenced by our estimation of their precision and by how much confidence we have in our predictive models. When confidence in our models is high, prediction errors are broadly suppressed, with only the narrowly defined and highly precise prediction errors that are relevant to the model being attended to – as occurs with tightly controlled physiological variables that are processed by reflex arcs.

When the confidence in our models is lower, as might occur in ambiguous and uncertain social environments, prediction errors are upweighted (even if not having high precision), and without being able to be explained at lower levels, the prediction errors enter conscious awareness as affective experiences that engender behavioural changes to resolve them. As Freud (1957) puts it, affect is 'a measure of the demand made upon the mind for work in consequence of its connection with the body'.

The precision weighting of prediction errors – and the balance between top-down prediction error and bottom-up prediction error – is maintained by neuromodulatory systems. Our major neuromodulators are the monoamines (dopamine, serotonin, noradrenaline, etc.) which act to adjust the excitatory–inhibitory balance enacted by glutamatergic and GABAergic neurons. These neuromodulators arise from brainstem and midbrain nuclei, and are under the top-down control of cortical networks that coordinate precision weighting in the overall interests of allostasis.

Symptoms of depression as precision-weighted prediction error

Bodily affects, such as hunger, thirst and pain, and emotions, such as fear, anger and joy, are brought to conscious awareness via short-term dynamical changes in precision weighting. Moods – which are more enduring affective states – are established by the long-term averages of precision weighting. The establishment of revised setpoints for precision weighting – which determine the mood state – reflects changes in model confidence. Depression is said to arise when lower-level models reflect an increase in uncertainty, and higher-level models encode a certainty about that uncertainty (Clark et al., 2018).

The anticipated uncertainty is reflected in upweighted interoceptive prediction errors. The higher-order confidence in this uncertainty then becomes self-reinforcing: the low-precision upweighted prediction errors fail to update the models, and there is a low expectation of reward, with reward-related prediction errors being attenuated and failing to update the expectations (Clark et al., 2018). The brain becomes 'locked in' (Barrett et al., 2016): the depressed person stops exploring the world and becomes increasingly internally preoccupied.

The altered confidence in the generative models is expressed in changes at the synapse, where the neuromodulators that weight precision exert their influence. It is for this that Friston (2023) describes mental illnesses such as depression as *synaptopathies*. But to describe depression in this way is not to side with the influence of biology against social effects. The active inference framework provides a way of linking social determinants to brain function. Poverty and marginalisation create instability, which are reflected in generative models that encode the uncertainty at the synaptic level by changes in the precision weighting of prediction errors (Badcock et al., 2017). Social determinants are sometimes discussed as causing depression as if they travel through an aether or miasma: but they must ultimately affect brain function to cause depression, and the active inference model outlines a mechanistic framework by which this might occur.

The framework helps to explain individual differences in the types of symptoms that are experienced. The pattern of symptoms depends on how the generative models are affected: on how a person's prior experiences interact with their social circumstances to influence the generative models and the prediction errors that manifest in the symptoms of depression. Increased uncertainty and the upweighting of interoceptive prediction errors are associated with anxiety-related symptoms that arise from autonomic reactivity, along with pain sensitivity, gastrointestinal symptoms and fatigue. The relationship with fatigue and sleep breaks down in depression: either the fatigue does not lead to restorative sleep (there is insomnia) or excessive sleep does not resolve the fatigue.

Reduced expectation of reward entails symptoms such as anhedonia, hopelessness, social withdrawal, reduced appetite and lack of sexual interest. Model rigidity (the locked-in brain) explains poor concentration and internal preoccupation, manifested as rumination. The pattern of symptoms will vary according to how predictive models are perturbed and provides a framework for understanding how symptomatic patterns can be so diverse. While the symptoms might vary, poor social functioning is ubiquitous. From this perspective, depression can be viewed as a disorder of

allostasis: as a failure to manage the body in the interests of its adaptation to the social environment (Badcock et al., 2017; Barrett et al., 2016).

Conclusion

Friston's active inference framework provides a means of narrowing Jaspers' explanation–understanding gap. It provides a plausible account for how symptoms arise in the context of the brain's relationship with the body and the psychosocial environment. It lays out a pathway for a new nosology, where the focus is on symptoms that cluster together, but in patterns that are not consistent with our current classification. Two new frameworks have been developed to better account for psychopathology: the Research Domain Criteria (RDoC) and the Hierarchical Taxonomy of Psychopathology (HiTOP). They have yet to be adopted to a significant extent in clinical settings, and as research tools lack coherent empirical frameworks. Perhaps, the active inference framework can be used to inform a more principled coherence.

There is much work to do to create a clinically useful nosological system. The active inference account of depression can seem abstruse. It derives from a mathematical formalism with respect to biological systems, which has then been applied secondarily to explain clinical phenomena. Interoceptive and affective experiences do not easily lend themselves to empirical investigation, but the active inference framework needs to put forward empirically testable hypotheses that can be disproved.

Active inference, nonetheless, provides a compelling model of brain function, and how it is affected by depression. Psychiatry, says Jaspers, 'is impelled to make use of methods that have been perfected elsewhere in order to improve the status of its subject matter, which is unique and irreplaceable for our apprehension of the world and humanity'. Our subject matter is important: we need to better explain depression so that we can improve our treatments, and in understanding the nature of the experiences of people with depression, we learn more about what it is to be human.

Author's note

This is the manuscript form of a presentation delivered for the 86th Beattie Smith Lecture on 16 November 2023. The lecture has been held since 1925 at the University of Melbourne on topics related to mental illness.

Acknowledgements

The author is grateful to Paul Badcock and Ben Harrison for their insightful feedback on the content of the presentation and manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Christopher G Davey https://orcid.org/0000-0003-1431-3852

References

Badcock PB, Davey CG, Whittle S, et al. (2017) The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences* 21: 182–194.

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

Borsboom D (2017) A network theory of mental disorders. *World Psychiatry* 16: 5–13.

Clark JE, Watson S and Friston KJ (2018) What is mood? A computational perspective. *Psychological Medicine* 48: 2277–2284.

Freud S (1957) Instincts and their vicissitudes. In: Freud S (ed.) *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 14. London: Hogarth Press, pp. 109–140.

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

Friston K (2023) Computational psychiatry: From synapses to sentience. *Molecular Psychiatry* 28: 256–268.

Ghaemi SN (2013) Understanding mood disorders: Karl Jaspers' biological existentialism. In: Stanghellini G and Fuchs T (eds) *One Century of Karl Jaspers' General Psychopathology*. Oxford: Oxford University Press, pp. 258–275.

Hacking I (2013) Lost in the forest. *London Review of Books* 35: 7–8.

Jaspers K (1959) *General Psychopathology*. Baltimore, MD: Johns Hopkins University Press.

Shorter E (2007) The doctrine of the two depressions in historical perspective. *Acta Psychiatrica Scandinavica* s443: 5–13.