The Separation Distress Hypothesis: The Ultimate Theory of Depression?

*Paul B. Badcock[1,2] and Karl J. Friston[3]

[1] Centre for Youth Mental Health, The University of Melbourne, Melbourne, 3054, Australia
[2] Orygen, Melbourne, 3054, Australia.
[3] Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, London WC1N 3AR, UK.

*Correspondence: pbadcock@unimelb.edu.au

We would like to applaud Watt's (2023) eloquent and thoroughgoing treatment of the intimate connection between depression and the social world. His survey of the neurobiological, psychiatric, and psychological literature — in support of the separation distress thesis — is impressive. The author provides the reader with an in-depth and expansive review of the various ways in which insults that stem from the social environment are implicated in depressive outcomes. A convincing argument is made that it is (beyond) time for researchers and clinicians to question their longstanding bioreductionist commitments in favour of aetiological models and interventions that firmly situate individuals within their broader, interpersonal milieu. We likewise agree that insights gleaned from an evolutionary analysis cast clear light on the question of *why* people become depressed; especially given that certain aspects of the depressed phenotype are conserved phylogenetically.

Despite these points of agreement, one could press Watt on his answer to this last question. The idea that the majority of depressive presentations can be explained in terms of separation distress, not to mention psychodynamic constructs more broadly, is open to challenge. First, there is the risk of infantilising depressive reactions by extrapolating a response to attachment dynamics — observed in infancy — to explain depressive outcomes later in life. *Prima facie*, given that the peak period of first onset for depressive illness occurs during the transition from adolescence to adulthood (Solmi et al., 2022), depression appears to be a largely adult phenomenon, suggesting that it has more to do with problems navigating the uncertainties of the social world after one has left the nest (Davey et al., 2008). Of course, this does not mean that early insults in the social environment are unrelated to depression — the evidence behind the deleterious effects of childhood abuse and neglect is overwhelming — but these can just as

readily be understood as vulnerability factors for later depressive onset, characterised by an increased susceptibility to social stress, without requiring recourse to separation distress *per se*.

This brings us to the more poignant question of whether one should privilege the author's account over other evolutionary views. Traditionally, theorising in this area has fallen into two main camps: the attachment school, based largely on the work of Bowlby (2008), which emphasises the loss of social connections in depression; and the social competition view, which focuses on loss of status and the failure of acquisitive behaviours (Gilbert, 1997; Price et al., 1994). The author's hypothesis belongs firmly in the former camp, and his extension of attachment theory to explain our responses to social defeat or humiliation later in life is extrapolative. Of course, other evolutionary proposals have also been raised, which attempt to provide a more unified, socially-oriented explanation for depression. Widely-cited examples include the social navigation hypothesis (Watson & Andrews, 2002), the social risk hypothesis (Allen & Badcock, 2003), and the social signal transduction theory (Slavich & Irwin, 2014). The implicit challenge is then to find evidence or predictions that disambiguate among evolutionary theories, and to determine whether the respective theories bring unique insights to the table.

Much of the author's review also concentrates on literature surrounding depressive disorder, which many have argued is a questionable target for evolutionary analysis (Nesse, 2000; Nettle, 2004). Rather, adaptive traits are likely to be species-typical, suggesting a need to concentrate on the mild-to-moderate, transient states of depression that we all experience from time to time (Allen & Badcock, 2006). In so doing, one is obliged to specify the adaptive functions of such states. Although the author suggests that depression serves to terminate protracted separation distress, he does not go on to describe, in mechanistic detail, how the depressive response achieves this. As we discuss below, these mechanisms have been a recent focus in computational psychiatry.

Such caveats aside, there are many aspects of the separation distress model that resonate with the evidence-base and have important implications for practice. The author also convincingly demonstrates that many of its predictions are born out in the literature. However, separation distress is likely to be only part of a larger story. Given the heterogeneity of depression, evolutionary theorists require a biologically plausible and integrative approach that is capable of explaining the full diversity of depressed states, while still remaining open to explanatory pluralism. To this end, we briefly consider recent advances stemming from theoretical psychiatry.

One potentially useful narrative — that speaks to separation distress — emerges from recent developments in computational psychiatry; in particular, first principle accounts of sentient behaviour based upon active inference (Schwartenbeck & Friston, 2016; Stephan et al. 2016; Smith et al. 2021; Adams et al. 2022). The arguments here are straightforward. If one can understand psychology in terms of (planning as) inference (Attias 2003; Botvinick & Toussaint 2012; Parr & Friston 2018; Linson, Parr, & Friston 2020), then psychopathology should yield to an explanation in terms of false inference — for example, inferring something is not there when it is (e.g., dissociative symptoms and neglect syndromes) or inferring something is there, when it is not (e.g., hallucinations and delusions). From a technical perspective, the phylogenetic conservation of our capacity for depressed states of mind suggests that this capacity enhances adaptive fitness or the marginal likelihood of the accompanying phenotypic trait: i.e., it is somehow [Bayes] optimal in the sense of natural (Bayesian model) selection or self-evidencing (Hohwy 2016; Badcock et al. 2017; Badcock, Friston, & Ramstead 2019). In turn, this leads to the notion that pathological depression is associated with aberrant prior beliefs that underwrite inference and choice behaviour. So, what is the nature of these aberrant priors?

Commented [MZ1]: I may be misunderstanding, but if the inference is being made by the subject, shouldn't this be reversed? (e.g., a hallucination involves inferring something is there when it is not). If this indeed is what you meant, then no revision is necessary; just checking.

Commented [PB2R1]: Thanks for your keen eye, Maggie – you're absolutely right.

When applying active inference to account for depressed mood and negatively valenced emotional states, one is drawn to inferences about the uncertainty, volatility, or reliability of the lived world. In the technical literature, this is usually framed in terms of the precision (i.e., confidence or reliability) afforded to various beliefs or sources of sensory (or indeed prosocial) evidence (Seth and Friston 2016; Smith et al. 2020; Friston 2022). In terms of the accompanying mechanistic or process theories, these kinds of inferences are thought to be manifest in terms of neuromodulation; increasing the gain of certain neuronal message passing, when it conveys more precise information (Adams et al. 2013; Edwards et al. 2012). Psychologically, this corresponds to the selection — in the sense of selective attention or sensory attenuation — of evidence for one's (e.g., subpersonal, Bayesian) beliefs.

Negatively valenced states are usually associated with increases in uncertainty (Linson, Parr, & Friston 2020; Badcock et al. 2017; Peters, McEwen, & Friston 2017; Lyndon & Corlett 2020). This is interesting from two perspectives. First, there is evidence to suggest that episodic fluctuations in emotional state entail revisions in the precision of various beliefs in cortical and subcortical hierarchies, while persistent low mood and chronic anxiety reflects an impoverished revision of beliefs about the precision of precision (Peters, McEwen, & Friston 2017; Clark,

Watson, & Friston 2018). This speaks to the importance of early life events in establishing priors over the predictability of the (prosocial) world (Cittern et al. 2018) and a potential predisposition to aberrant emotional responses to changes in circumstances (e.g., life events such as bereavement). This foregrounds the putative role of separation distress; as a Bayes optimal response to an increase in uncertainty about the interpersonal, prosocial, and encultured world – and how a suboptimal Bayesian prior or bias could confound the resolution of this uncertainty. Second, from a mechanistic point of view, the mediation of precision in the central nervous system and beyond (e.g., neuroendocrine and immune systems) implicates exactly the modulatory systems surveyed in the target article (Friston 2022; Bhat et al. 2021; Peters, McEwen, & Friston, 2017).

Although one could expand on related active inference accounts of emotion and depression (e.g., Barrett & Simmons, 2015; Chekroud, 2015; Joffily & Coricelli, 2013; Smith et al., 2019; Smith, Parr, & Friston, 2019; Smith et al., 2021), we would like to concentrate on our evolutionary systems theory, which combines active inference with research spanning psychology, psychiatry, and neuroscience, to emphasise the crucial role of social context in depressive phenomena (see Badcock et al., 2017). According to this view, normative levels of depressed mood instantiate an adaptive, socially risk-averse strategy that functions to reduce socioenvironmental volatility when sensory cues indicate an increased likelihood of unexpected or non-preferred social outcomes (e.g., rejection, defeat, or loss). The depressive response achieves this function by causing changes in perception (e.g., an increase in the precision afforded to incoming social stimuli); suppressing confident, reward-approach behaviours (e.g., anhedonia and social withdrawal); and generating signalling behaviours that either elicit support (e.g., reassurance-seeking) or defuse conflict (e.g., submissive behaviours). Notably, recent work in computational psychiatry has applied active inference to show how this adaptive response successfully plays out *in silico* but can also lead to maladaptive depressive outcomes (Constant et al., 2022).

On this point, in most cases, the depressive response functions adaptively by attracting interpersonal support and reducing social uncertainty through risk-averse interpersonal behaviours and rapid belief-updating when faced with unexpected social outcomes. However, psychopathology can emerge when there are ongoing discrepancies between actual and preferred social outcomes over time (i.e., chronic prediction errors). Such chronic social stress can entrench aberrant prior beliefs that social rewards are unlikely (e.g., pessimism, low self-worth, shame), which perpetuate risk-averse depressive behaviours (e.g., social withdrawal) and lead to

dysfunctional responses (e.g., learned helplessness; see Chekroud, 2015; Kube et al., 2020). Consistent with the author's view, this model suggests that vulnerability to psychopathology often arises from early exposure to social stress (e.g., parental abuse or neglect), which promotes prior beliefs that social outcomes are uncontrollable and heightens the sensitivity of stress response systems to interpersonal stressors (e.g., inflammatory immune responses). On the other hand, the model also recognises that clinical depression can emerge from *asocial* causes that produce neurobiological abnormalities implicated in depressed mood (such as proinflammatory immune responses induced by chronic pain or physical illness; see, for example, Bhat et al., 2021).

In brief, the (mechanistic) model discussed in this commentary suggests that depression is an adaptive response that actively reduces exposure to surprising or deleterious social outcomes. Under this view, separation distress might be seen as but one of a suite of strategies that have been favoured by selection to return the individual to a state of social homeostasis. One of the benefits of this perspective is that it readily incorporates the author's hypothesis — not to mention the wide-ranging evidence he calls upon to support it — without discounting insights drawn from other evolutionary models. Much like the author's hypothesis, the evolutionary systems theory of depression yields clear implications for prevention and intervention efforts by underscoring the importance of improving individuals' social contexts (e.g., interpersonal psychotherapy; see Cuijpers et al., 2011). Mechanistically, an active inference formulation is also accompanied by promising conceptual, computational, and empirical tools that have begun to cast new light on psychopathology (Smith et al., 2021), not to mention the mind, brain and behaviour (see Parr et al., 2022).

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry*, *4*, 47. https://doi.org/10.3389/fpsyt.2013.00047

Adams, R. A., Vincent, P., Benrimoh, D., Friston, K. J., & Parr, T. (2022). Everything is connected: inference and attractors in delusions. *Schizophrenia Research*, *245*, 5-22. https://doi.org/10.1016/j.schres.2021.07.032

Attias, H. (2003, January). Planning by probabilistic inference. In *Proceedings of the 9th international workshop on artificial intelligence and statistics* (pp. 9-16). PMLR.

Allen, N. B., & Badcock, P. B. (2003). The social risk hypothesis of depressed mood: evolutionary, psychosocial, and neurobiological perspectives. *Psychological Bulletin, 129*(6), 887. https://doi.org/10.1037/0033-2909.129.6.887

Allen, N. B., & Badcock, P. B. (2006). Darwinian models of depression: A review of evolutionary accounts of mood and mood disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 30*(5), 815-826. https://doi.org/10.1016/j.pnpbp.2006.01.007

Badcock, P. B., Friston, K. J., & Ramstead, M. J. (2019). The hierarchically mechanistic mind: A free-energy formulation of the human psyche. *Physics of Life Reviews*, *31*, 104-121. https://doi.org/10.1016/j.plrev.2018.10.002

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: an evolutionary systems theory. *Trends in Cognitive Sciences, 21*(3), 182-194. https://doi.org/10.1016/j.tics.2017.01.005

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16*(7), 419-429. https://doi.org/10.1038/nrn3950

Bhat, A., Parr, T., Ramstead, M., & Friston, K. (2021). Immunoceptive inference: why are psychiatric disorders and immune responses intertwined? *Biology & Philosophy*, *36*(3), 27. https://doi.org/10.1007/s10539-021-09801-6

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485-488. https://doi.org/10.1016/j.tics.2012.08.006

Bowlby, E. (2008). *Loss-Sadness and Depression: Attachment and Loss Volume 3* (Vol. 3): Random House.

Chekroud, A. M. (2015). Unifying treatments for depression: an application of the Free Energy Principle. *Frontiers in Psychology, 6*, 153. https://doi.org/10.3389/fpsyg.2015.00153

Cittern, D., Nolte, T., Friston, K., & Edalat, A. (2018). Intrinsic and extrinsic motivators of attachment under active inference. *PloS One*, *13*(4), e0193955. https://doi.org/10.1371%2Fjournal.pone.0193955

Clark, J. E., Watson, S., & Friston, K. J. (2018). What is mood? A computational perspective. *Psychological Medicine*, *48*(14), 2277-2284. https://doi.org/10.1017%2FS0033291718000430

Constant, A., Hesp, C., Davey, C. G., Friston, K. J., & Badcock, P. B. (2021). Why depressed mood is adaptive: a numerical proof of principle for an evolutionary systems theory of depression. *Computational Psychiatry, 5*(1), 60. https://doi.org/10.5334%2Fcpsy.70

Cuijpers, P., Geraedts, A. S., van Oppen, P., Andersson, G., Markowitz, J. C., & van Straten, A. (2011). Interpersonal psychotherapy for depression: a meta-analysis. *American Journal of Psychiatry, 168*(6), 581-592. https://doi.org/10.1176/appi.ajp.2010.10101411

Davey, C. G., Yücel, M., & Allen, N. B. (2008). The emergence of depression in adolescence: Development of the prefrontal cortex and the representation of reward. *Neuroscience & Biobehavioral Reviews, 32*(1), 1-19. https://doi.org/10.1016/j.neubiorev.2007.04.016

Edwards, M. J., Adams, R. A., Brown, H., Parees, I., & Friston, K. J. (2012). A Bayesian account of 'hysteria'. *Brain*, *135*(11), 3495-3512. https://doi.org/10.1093/brain/aws129

Friston, K. (2023). Computational psychiatry: from synapses to sentience. *Molecular Psychiatry, 28*(1), 256-268. https://doi.org/10.1038/s41380-022-01743-z

Gilbert, P. (1997). The evolution of social attractiveness and its role in shame, humiliation, guilt and therapy. *British Journal of Medical Psychology, 70*(2), 113-147. https://doi.org/10.1111/j.2044-8341.1997.tb01893.x

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259-285. https://doi.org/10.1111/nous.12062

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Computational Biology, 9*(6), e1003094. https://doi.org/10.1371/journal.pcbi.1003094

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry, 87*(5), 388-398. https://doi.org/10.1016/j.biopsych.2019.07.017

Linson, A., Parr, T., & Friston, K. J. (2020). Active inference, stressors, and psychological trauma: A neuroethological model of (mal) adaptive explore-exploit dynamics in ecological context. *Behavioural Brain Research*, *380*, 112421. https://doi.org/10.1016/j.bbr.2019.112421

Lyndon, S., & Corlett, P. R. (2020). Hallucinations in posttraumatic stress disorder: Insights from predictive coding. *Journal of Abnormal Psychology*, *129*(6), 534-543. https://psycnet.apa.org/doi/10.1037/abn0000531

Nesse, R. M. (2000). Is depression an adaptation? *Archives of General Psychiatry, 57*(1), 14-20. https://doi.org/10.1001/archpsyc.57.1.14

Nettle, D. (2004). Evolutionary origins of depression: a review and reformulation. *Journal of Affective Disorders, 81*(2), 91-102. https://doi.org/10.1016/j.jad.2003.08.009

Parr, T., & Friston, K. J. (2018). The anatomy of inference: generative models and brain structure. *Frontiers in Computational Neuroscience*, *12*, 90. https://doi.org/10.3389/fncom.2018.00090

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: the free energy principle in mind, brain, and behavior*: MIT Press.

Peters, A., McEwen, B. S., & Friston, K. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology*, *156*, 164-188. https://doi.org/10.1016/j.pneurobio.2017.05.004

Price, J., Sloman, L., Gardner, R., Gilbert, P., & Rohde, P. (1994). The social competition hypothesis of depression. *The British Journal of Psychiatry, 164*(3), 309-315. https://doi.org/10.1192/bjp.164.3.309

Schwartenbeck, P., & Friston, K. (2016). Computational phenotyping in psychiatry: a worked example. *eNeuro*, *3*(4), 0049-16.2016. https://doi.org/10.1523/eneuro.0049-16.2016

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

Slavich, G. M., & Irwin, M. R. (2014). From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychological Bulletin, 140*(3), 774. https://doi.org/10.1037/a0035302

Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences, 75*(1), 3-13. https://doi.org/10.1111/pcn.13138

Smith, R., Parr, T., & Friston, K. J. (2019). Simulating emotions: An active inference model of emotional state inference and emotion concept learning. *Frontiers in Psychology*, *10*, 2844. https://doi.org/10.3389/fpsyg.2019.02844

Smith, R., Lane, R. D., Parr, T., & Friston, K. J. (2019). Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their

potential clinical relevance. *Neuroscience & Biobehavioral Reviews*, *107*, 473-491. https://doi.org/10.1016/j.neubiorev.2019.09.002

Smith, R., Kirlic, N., Stewart, J. L., Touthang, J., Kuplicki, R., Khalsa, S. S., ... & Aupperle, R. L. (2021). Greater decision uncertainty characterizes a transdiagnostic patient sample during approach-avoidance conflict: a computational modelling approach. *Journal of Psychiatry and Neuroscience*, *46*(1), E74-E87. https://doi.org/10.1503/jpn.200032

Smith, R., Kuplicki, R., Feinstein, J., Forthman, K. L., Stewart, J. L., Paulus, M. P., ... & Khalsa, S. S. (2020). A Bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *PLoS Computational Biology*, *16*(12), e1008484. https://doi.org/10.1371/journal.pcbi.1008484

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., . . . Kim, J. H. (2022). Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry, 27*(1), 281-295. https://doi.org/10.1038/s41380-021-01161-7

Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., ... & Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, *10*, 550. https://doi.org/10.3389/fnhum.2016.00550

Watson, P. J., & Andrews, P. W. (2002). Toward a revised evolutionary adaptationist analysis of depression: The social navigation hypothesis. *Journal of Affective Disorders, 72*(1), 1-14. https://doi.org/10.1016/s0165-0327(01)00459-1

Watt, D.F. (2023). [editors to add reference when submitting into production]