

From Broken Models to Treatment Selection: Active Inference as a Tool to Guide Clinical Research and Practice

Lukas Kirchner¹, Anna-Lena Eckert², Max Berg¹

[1] *Department of Psychology, Clinical Psychology and Psychotherapy, Philipps-University of Marburg, Marburg, Germany.* [2] *Department of Psychology, Theoretical Cognitive Science, Philipps-University of Marburg, Marburg, Germany.*

Clinical Psychology in Europe, 2022, Vol. 4(2), Article e9697, <https://doi.org/10.32872/cpe.9697>

Published (VoR): 2022-06-30

Corresponding Author: Lukas Kirchner, Department of Psychology, Clinical Psychology and Psychotherapy, Philipps-University of Marburg, Gutenbergstraße 18, 35037 Marburg, Germany. E-mail: lukas.kirchner@uni-marburg.de

Computational theories have fundamentally changed the scientific understanding of how the mind works for both healthy and pathological experiences and behaviours. In this context, the active inference framework has gained considerable attention within the scientific community (Heins et al., 2022; Smith et al., 2022). As a process theory, it integrates complex phenomena, such as perception, learning, and action under a unified theory of Bayesian inference (Da Costa et al., 2020; Friston et al., 2017). Active inference has proven useful in modelling data from heterogeneous fields ranging from cognitive neuroscience to biology and general psychology (e.g., Friston et al., 2016, 2017). Its computational tractability and biological plausibility have also made it increasingly relevant to clinical psychology in recent years (e.g., Smith, Badcock, et al., 2021).

In active inference and related, Bayesian neurocomputational theories, it is assumed that individuals do not have direct access to the circumstances in their surroundings. Instead, they have to infer the (probabilistic) properties of their environment through action and perception by integrating prior information about their environment with ambiguous sensory input in a rational (i.e., Bayes-optimal) manner (Friston et al., 2016; Hohwy et al., 2008). The resulting “internal model of the world” (i.e., the agent’s beliefs about how certain sensory information relates to environmental conditions) shapes future perception (Friston, 2010) and enables agents to leverage the past to predict the future in an ever-changing environment (Badcock et al., 2017). In accordance with this perspective, perception, action, and learning are all subject to inferential process-



es on different timescales (Smith et al., 2022). Individuals thus take an active role in constructing their experiences, which they can further alter by actively changing their environment.

Whereas accurate internal models provide good predictions about which action sequences lead to preferred sensory observations, distorted internal models can lead to aberrant experiences and behaviours that may hinder the organism from achieving its goals (Badcock et al., 2017; Schwartenbeck et al., 2015). There is evidence that the internal models of individuals with mental disorders show substantial deviations from each other and from the models of healthy individuals. For example, recent accounts of depression have conceptualised patients' tendency to reappraise or disregard positive information in terms of highly precise and hence tenacious negative prior beliefs (Kube et al., 2020). This computational perspective has inspired novel ideas for the treatment of depression (e.g., Cherkroud, 2015). Similar reconceptualisations with relevance for psychological treatments have been suggested for numerous mental disorders and health conditions, including psychosis (Sterzer et al., 2018), persistent somatic symptoms (Paulus et al., 2019), and eating disorders (Barca & Pezzulo, 2020).

In this context, the active inference framework offers the opportunity to formalise deviations in a person's internal models, thus enabling a detailed description and operationalisation of relevant experiential and behavioural distortions (e.g., Montague et al., 2012). From our point of view, this could improve clinical research, diagnostics, and the treatment of mental disorders in several ways. Clinical research may benefit from a finely grained formalisation of deviant experiences and behaviours within the active inference framework because it opens up a possibility for studying aetiological mechanisms from a computational perspective (Stephan, Binder, et al., 2016). Because of their generative structure, computational theories enable the derivation of well-operationalised hypotheses about pathological processes in mental disorders and the computer simulation of aberrant experience and behaviour. In comparison with empirical data, researchers could thus rigorously formalise, simulate, and compare different mechanistic models of patients' experiential and behavioural symptoms. Moreover, the active inference perspective could inform the diagnostics of mental disorders (or rather the diagnostics of patients' implicit belief systems) by providing practitioners with individual estimates of their patients' internal model parameters in disorder-relevant situations (Stephan, Bach, et al., 2016). For example, using probabilistic gambling tasks that distinguish between goal-directed information seeking and random exploration behaviour could provide clinicians with individual parameter diagnostics regarding the relationship between information seeking, reward sensitivity, and psychopathology in substance abusers (Smith, Schwartenbeck, et al., 2021). Such applications could not only strengthen a more transdiagnostic perspective on mental disorders, but also have tangible implications for treatment development and treatment selection. If we assume that therapeutic interventions may have different effects on patients' model parameters, finely grained

operationalisation in the context of active inference could contribute to tailored interventions that target specifically these parameters. From the practitioner's perspective, it would be particularly important to investigate which tasks are likely to diagnose internal models that inform treatment selection and guide psychotherapy.

The active inference approach affords an improved mechanistic understanding of pathological processes in mental disorders. As a unifying process theory of brain and mind function, it brings together perception, learning, action, and decision making under the umbrella of a Bayesian principle, which predestines it for clinical application. Because of its high degree of formalisation and its flexibility, we believe that the active inference approach is well suited to functionally link heterogeneous clinical phenomena to patients' internal belief systems. This will enable researchers to better differentiate and operationalise underlying mechanisms and tailor the diagnosis, aetiology, and treatment of mental disorders.

Funding: This work was funded by the Hessian Ministry of Higher Education, Research, Science, and the Arts.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: This work was realised within the cluster initiative "The Adaptive Mind" (TAM) which brings together scientists from experimental psychology, clinical psychology, and artificial intelligence to improve the understanding of how the human mind successfully adapts to changing conditions. TAM is funded by the Hessian Ministry of Higher Education, Research, Science, and the Arts.

References

- 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>
- Barca, L., & Pezzulo, G. (2020). Keep your interoceptive streams under control: An active inference perspective on anorexia nervosa. *Cognitive, Affective & Behavioral Neuroscience*, 20(2), 427–440. <https://doi.org/10.3758/s13415-020-00777-6>
- Chekroud, A. M. (2015). Unifying treatments for depression: An application of the Free Energy Principle. *Frontiers in Psychology*, 6, Article 153. <https://doi.org/10.3389/fpsyg.2015.00153>
- Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99, Article 102447. <https://doi.org/10.1016/j.jmp.2020.102447>
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>

- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience and Biobehavioral Reviews*, 68, 862–879. <https://doi.org/10.1016/j.neubiorev.2016.06.022>
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49. https://doi.org/10.1162/NECO_a_00912
- Heins, C., Millidge, B., Demekas, D., Klein, B., Friston, K., Couzin, I. D., & Tschantz, A. (2022). pymdp: A Python library for active inference in discrete state spaces. *Journal of Open Source Software*, 7(73), Article 4098. <https://doi.org/10.21105/joss.04098>
- Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3), 687–701. <https://doi.org/10.1016/j.cognition.2008.05.010>
- 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>
- Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1), 72–80. <https://doi.org/10.1016/j.tics.2011.11.018>
- Paulus, M. P., Feinstein, J. S., & Khalsa, S. S. (2019). An active inference approach to interoceptive psychopathology. *Annual Review of Clinical Psychology*, 15(1), 97–122. <https://doi.org/10.1146/annurev-clinpsy-050718-095617>
- Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Medical Hypotheses*, 84(2), 109–117. <https://doi.org/10.1016/j.mehy.2014.12.007>
- 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., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, 107, Article 102632. <https://doi.org/10.1016/j.jmp.2021.102632>
- Smith, R., Schwartenbeck, P., Stewart, J., Kulpicki, R., Ekhtiari, H., & Paulus, M. (2021). Reduced action precision and biased learning in substance users when solving the explore-exploit dilemma: An active inference modelling approach. *Biological Psychiatry*, 89(9), S55–S56. <https://doi.org/10.1016/j.biopsych.2021.02.155>
- Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., Heinz, A., Huys, Q. J. M., Owen, M. J., Binder, E. B., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Montague, P. R., Schnyder, U., Wang, X.-J., & Breakspear, M. (2016). Charting the landscape of priority problems in psychiatry, part 1: Classification and diagnosis. *The Lancet Psychiatry*, 3(1), 77–83. [https://doi.org/10.1016/S2215-0366\(15\)00361-2](https://doi.org/10.1016/S2215-0366(15)00361-2)
- Stephan, K. E., Binder, E. B., Breakspear, M., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Schnyder, U., Wang, X.-J., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Heinz, A., Huys, Q. J. M., Montague, P. R., Owen, M. J., & Friston, K. J. (2016). Charting the landscape of priority

problems in psychiatry, part 2: Pathogenesis and aetiology. *The Lancet Psychiatry*, 3(1), 84–90.

[https://doi.org/10.1016/S2215-0366\(15\)00360-0](https://doi.org/10.1016/S2215-0366(15)00360-0)

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, 84(9), 634–643. <https://doi.org/10.1016/j.biopsych.2018.05.015>

EACLIPT

Clinical Psychology in Europe (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.