

Making life & mind as clear as possible, but not clearer

Alex Gomez-Marin

Instituto de Neurociencias (CSIC-UMH), 03550 San Juan Alicante, Spain

agomezmarin@gmail.com <https://behavior-of-organisms.org/> @behaviOrganisms

Behavioral and Brain Sciences commentary on Bruineberg et al. (2021) "The Emperor's New Markov Blankets"

Abstract. Neuroscience needs theory. Ideas without data are blind, and yet mechanisms without concepts are empty. Friston's free energy principle paradigmatically illustrates the power and pitfalls of current theoretical biology. Mighty metaphors, turned into mathematical models, can become mindless metaphysics. Then, seeking to understand everything in principle, we may explain nothing in practice. Life can't live in a map.

*

In their brilliant but rather undiscovered book, *The Dialectical Biologist*, Richard Levins and Richard Lewontin write: "a theory that can explain everything explains nothing" (1985, p. 65). The paragraph is worth quoting in its entirety, but on another occasion. They refer to the ideological excesses of Marxism, Freudianism and Darwinism. In the context of 21st century life and mind sciences, a similar cautionary note could be made about Fristonism.

Already well-known for his contributions to functional imaging, the British neuroscientist Karl Friston has distinguished himself over the last decade as a steady proponent of the so-called free energy principle (FEP) (Friston, 2010). Friston borrows foundational concepts from thermodynamics and combines them with two good old ideas: Hermann von Helmholtz's understanding of perception as inference (1867) and Claude Bernard's insight on the self-regulatory nature of vital processes (1878). The FEP is a normative theory, and its main commandment is the minimization of surprise. Framed in a modern Bayesian framework –whereby beliefs are updated by means of loops of prediction, error, and correction–, the brain would be an active inference machine doing predictive processing.

But this is precisely what begs the question. It is not uncommon that a formal approach used by brain scientists to study other brains ends up surreptitiously postulated as what all brains actually do (Gomez-Marin, 2019): if *one can* make sense of brains as inference machines, then *brains must* be inference machines. Indeed, a pervasive temptation amongst scientists is to treat the conjunction "as if" as "is" (for example, animals are (like) machines, brains are (like) computers). We know that maps are not territories, but we still conflate them (Andrews, 2021).

The FEP illustrates what I call "the 3M fallacy", a sleight of mind whereby captivating Metaphors are turned into mathematical Models that then become covert or self-ignorant Metaphysics. Bruineberg et al. (2021) thus offer a timely dissection of the notion of Markov blankets as interpreted by some FEP proponents, making a distinction between mathematical tools to study organisms and ontological pronouncements about those very organisms. The critique would not be so patently urgent had Friston and collaborators refrained from elevating their theoretical framework from a powerful heuristic into a sort of biological theory of everything.

Having started as an ambitious integrative theory of brain function, the FEP has been progressively expanded to behavior, the mind, and even consciousness! Perception, decision-making and learning would all be explainable under the FEP too, in animals but also in plants and microorganisms. And yet, one should be hesitant of any sweeping principle that accounts for nematode foraging and romantic love. Seeking an understanding of virtually everything in principle, the framework risks explaining actually nothing in practice.

Is Friston's FEP neurobiology's "string theory"? The latter has been deemed by prominent theoretical physicists as "not even wrong" (Woit, 2006). The FEP is no less ambitious. Sufficiently elaborated for professional acolytes to revere and pursue, it appears sufficiently vague to be immune to empirical data. Without indulging in falsification chauvinism, one should at least ask whether and how the FEP could be wrong.

Perhaps, like Darwin's, Friston's theory may not be disprovable by any one experiment, while still offering a powerful overall way to see things. Back to Freud and Marx, their hypotheses have also tremendously influenced modern thought, whether directly testable or not. Isn't it paradoxical that Darwin's account (1859), arguably the greatest of all biological theories (and firmly grounded in empirical observation), cannot be falsified by a single aberrant finding? Biology is the paradigmatic science of exceptions. Consistency is overrated; coherency is scarce. Overwhelmed by analytical facts, we desperately need synthetic views. Science accomplishes something remarkable

when, apart from discovering causal mechanisms, it offers organizing principles. A virtuous middle may be found where castles in the air meet castles made of sand.

Should theoreticians rejoice at mass-producing prêt-à-porter “conceptual blankets”, or instead strive to offer tailor-made suits to the phenomena of life? Concerned with clarity we lose touch with concrete reality, as the physicist’s joke of the spherical cow infamously illustrates. Pushing the blanket analogy a bit further, let us ask: how fit should the cloth be? A loose one can fit almost everything, but properly fit nothing. A tight one does a proper job, but just once. The “comfy blanket” may be a suitable garment to watch a movie at home, but a tremendously inappropriate one to attend a funeral. Paradoxically, as Iain McGilchrist remarks, “[k]nowledge of something that is by its nature not precise will *itself* have to be imprecise, if it is to be accurate” (2021, p. 583). Our duty as thinkers is to be as clear as possible, but not clearer.

In that respect, the FEP has remarkable ingredients. It puts action into perception, re-enacting Henri Bergson’s pioneering theory of perception as virtual action (1896). It also emphasizes the life-mind continuity (Thompson, 2010), but see (Bitbol & Gallagher, 2018). The framework, however, leaves crucial biological aspects out. Based on statistical averages in steady states, it struggles to take historicity and individuality into account. Moreover, it overlooks the developmental origin of Bayesian priors (Ciaunica et al., 2021). Evolutionarily, the FEP focuses on important similarities across species, but dispenses with the relevant differences, say, between a shiitake mushroom and a beluga whale.

In sum, we can appreciate the most fecund aspects of the theory without mistaking stupendous abstractions for concrete reality. Such is the Herculean challenge faced when seeking knowledge of life as we know it versus as it is lived (Canguilhem, 1952). The objectivist stance disregards the perspective of each living organism. “We think in generalities, but we live in detail”, wrote Alfred North Whitehead (1926, p.192). Theories of everything can hardly be theories of *every* thing, since nothing is anything except in its context. Conversely, a theory of every *thing* cannot be a theory of everything since, to our dismay (and sometimes denial), physics does not account for the existence of lived experience. Theory strives for simplicity, but life is complicated. Living organisms can’t live in a map.

Acknowledgements. I thank Michel Bitbol, Tim Elmo Feiten, Iain McGilchrist and Rupert Sheldrake for insightful discussions.

References

- Andrews, M. (2021) The math is not the territory: navigating the free energy principle. *Biology & Philosophy* 36:30.
- Bergson, H. (1896) *Matière et mémoire: Essai sur la relation du corps à l'esprit*. Paris.
- Bernard, C. (1878) *Leçons sur les phénomènes de la vie communes aux animaux et aux végétaux*. Paris, Baillière.
- Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2021) The Emperor’s New Markov Blankets. *Behavioral and Brain Sciences*, 1-63.
- Bitbol, M., Gallagher, S. (2018) The free energy principle and autopoiesis. *Phys Live Rev* 24:24-26.
- Canguilhem, G. (1952) *La Connaissance de la vie*. Paris: Vrin.
- Ciaunica, A., Constant, A., Preissl, H, Fotopoulou, K. (2021) The first prior: From co-embodiment to co-homeostasis in early life. *Consciousness and Cognition* 91: 103117.
- Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray, Albemarle Street.
- Friston, K. J. (2010). The free energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.
- Gomez-Marin, A. (2019) A clash of Umwelts: Anthropomorphism in behavioral neuroscience. *Behavioral and Brain Sciences*, 42, E229.
- von Helmholtz, H. (1867) *Handbuch der Physiologischen Optik*. Leipzig: Leopold Voss.
- McGilchrist, I. (2021) *The Matter with Things: Our Brains, our Delusions, and the Unmaking of the World*. Perspectiva.
- Levins, R., & Lewontin, R. (1985) *The Dialectical Biologist*. Harvard University Press.
- Thompson, E. (2010) *Mind in Life: Biology, Phenomenology and the Sciences of Mind*. Belknap Press, Harvard University Press.
- Whitehead, A. N. (1926) The Education of an Englishman. *The Atlantic Monthly* 138:192-8.
- Woit, P. (2006) *Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*. Basic Books.