The many faces of action: Comment on An active inference model of hierarchical action understanding, learning and imitation by Proietti, Pezzulo, and Tessari

By Karl Friston

The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London, UK WC1N 3AR and VERSES AI Research Lab, Los Angeles, CA 90016, USA

Email: k.friston@ucl.ac.uk

Keywords: *free energy principle; active inference; action observation; salience; novelty; affordance*

Proietti et al. [1] present numerical studies to delineate nearly every aspect of action; namely, its execution, observation, planning, understanding and imitation. This *tour de force* showcases every aspect of active inference and learning—and how these aspects fit together to reproduce the finer details of our enacted and embodied exchanges with the world (and others). Indeed, it was so comprehensive, I had to remind myself that all the phenomena demonstrated in [1] emerged from two simple imperatives: every representation is updated to minimise variational free energy—and every action is chosen to minimise expected free energy. In other words, every representation is updated to maximise the marginal likelihood of observations—and every action maximises expected information gain. With such simple rules, one might ask how the authors were able to explain diverse phenomena; ranging from action observation to imitation; from perceptual inference to learning; from saccadic eye movements to motor planning. Perhaps the simplest answer is that the art of applying active inference lies in specifying—or learning—the structure of the generative model apt for the explanatory targets at hand.

In short, if one gets the generative model right, everything else should follow for free. Everything else, in this instance, encompasses a number of fundamental cognitive and learning processes that are central to action observation and understanding. Getting the generative model right is just identifying the right functional architecture—and ensuring that this architecture can be plausibly mapped to functional brain anatomy. With this in mind, this commentary foregrounds a few key themes in [1].

"*Action understanding and execution rest on shared (ideomotor) codes across perception and action*" [1]

The notion of "shared ideomotor codes" is foundational here. It speaks to a view of motor control that active inference inherits from ideomotor theory [2, 3]. In short, we do not issue motor commands to produce desired movements; rather, we predict the sensory (e.g., proprioceptive) consequences of a movement we have committed to—and let our reflexes realise those predictions. This realisation is, literally, unremarkable (i.e., not sensed or perceived), unless our predictions are somehow thwarted, and we register our surprise that our intentions have not been realised.

Formally, under active inference, this emerges from the minimisation of variational free energy (i.e., proprioceptive prediction errors) by our embodied reflex arcs. In short, movement is all about realising predictions and plans under a suitable generative model. The generative (a.k.a., forward) model generates predictions of the consequences of action, such that one can read corollary discharge [4] as predictions of exteroceptive (e.g., visual) consequences of movement—a corollary to the accompanying predictions of proprioception engendering that movement. This brings us to the kernel of the work reported in [1]; namely action observation and imitation.

"*Hence, the functional role associated with the mirror neuron system is predictive action monitoring*." [1]

If we have generative (a.k.a. world) models that predict the consequences of action in multiple (exteroceptive, interoceptive and proprioceptive) domains, then exteroceptive (e.g., visual) predictions can explain observed actions, irrespective of whether they are generated by oneself or another. This means that mirror neuron phenomena are a necessary and emergent property of the generative models apt for explaining the kinds of actions executed by conspecifics [5-7]. The original active inference simulations of this [8] distinguished action execution from action observation by suspending proprioceptive sensations (by attenuating the precision of sensory prediction errors). Sensory attenuation subsequently proved a straightforward and fecund device to model communication; in which talking and listening (i.e., acting and observing) was implemented by switching between various states of (proprioceptive) sensory attenuation, under a shared generative model [9]. Initially, I was surprised that Pezzulo et al. did not pursue this approach. I then realised they were dealing with a more expressive generative model that entailed *two kinds of action*; namely, saccadic eye movements and limb movements, respectively. This was an ambitious move that enabled them to make a number of fundamental points about the structure of the requisite generative models and implicit functional anatomy.

"*Novelty entails a divergence from memory, while surprise entails a divergence from expectations*." [1]

The first point rests upon their careful treatment of the imperatives for action. Under active inference, the only imperative is to select actions that minimise expected free energy. This can be read in many ways. A particular decomposition of expected free energy into *risk* and *ambiguity* is provided in [1]. By rearranging the terms, one can also express expected free energy in terms of *expected information gain* and *expected value*, where value is scored by the log probability of preferred outcomes [10].

In this example, there were no explicit preferences, which meant that all behaviour was driven by the epistemic affordances implicit in the expected information gain (a.k.a., intrinsic value or motivation). This underwrites curiosity and intrinsic motivation in neurorobotics [11-14], and can itself be read from a number of perspectives. This follows from the fact that the expected information gain pertains to any random variable in the generative model; including states and parameters. The epistemic affordance due to expected information gain about latent states is generally referred to as *salience*. Similarly, the expected information gain associated with model parameters is referred to as *novelty* [15]. Interestingly, the expected information gain associated with parameters is the mutual information between latent causes and observable consequences. This is closely related to many objective functions in computational neuroscience and machine learning [16-18].

In the present setting, there is interesting dissociation between the two kinds of epistemic affordances (relating to states and parameters). This follows from considering eye and limb movements within the same generative model. The epistemic affordances for eye movements can be understood in terms of responding to salient cues that resolve uncertainty about states of affairs generating observations. Conversely, limb movements are driven by novelty to reduce uncertainty about the contingencies encoded by model parameters (here Dirichlet counts in various likelihood mappings within the model). This computational distinction is carefully unpacked in [1] in relation to the putative functional anatomy; asking where the requisite salience and novelty computations could be implemented in the brain. I thought that their treatment was a masterclass in trying to map the computational anatomy implied by a first principles account of inference and learning to the functional anatomy of the brain.

On my reading, the computation of salience is much easier to speculate about than the computation of novelty. For example, it is straightforward to imagine that the deep layers of the superior colliculus encode the expected information gain (a.k.a., expected Bayesian surprise) associated with the targets of putative saccades [19]. One wonders, given the review [1], whether a more anterior subcortical anatomy involving the cortico–basal ganglia–thalamic system might play a similar role in computing novelty. One knows—from the functional form of the requisite belief updates—that novelty is a function of particular likelihood mappings, which are often read in terms of extrinsic (between region) connections in the brain [20]. It would be interesting to pursue the analysis in [1] to work out the anatomical constraints on the computational of novelty that underwrites action selection.

"*While pursuing salience is useful to reduce the uncertainty about hidden states (or ambiguity), pursuing novelty is useful to reduce uncertainty about model parameters (or ignorance).*" [1]

There are many other issues it would be interesting to discuss. These include the separation of temporal scales that defines the hierarchical level in these kinds of generative models. One interesting twist—that the current model brings to the table—is the break from a linear hierarchy to produce a heterarchy that can integrate observation (with active vision) and motor execution; where both are informed by—and inform—a slower level of understanding, in terms of motor narratives and implicit intentions.

It may be useful to close with a generic point implicit in the above quote; namely, that reducing uncertainty about hidden states is necessary to reduce uncertainty about model parameters. This speaks to a general principle in active inference: active inference commits to a variational inversion of generative models that is operationally defined in terms of a mean field approximation or factorisation [21]. Intuitively, the generative model 'carves nature at its joints' to produce factors or modules that are independent, when conditioned upon higher levels of the model. In a similar way, there is an adiabatic factorisation, in which random variables are factorised into states and parameters; depending upon whether they change quickly or slowly. The point here is that updating Bayesian beliefs about states depends upon beliefs about parameters and updating beliefs about parameters depends on beliefs about states. In short, everything is contextualised by everything else.

Practically, this implies that the right kind of computational architecture has to include probabilistic representations of model parameters (i.e., connection strengths) for optimal learning. This begs the question of how this is done in the brain; in other words, how is the uncertainty about a particular synaptic connection encoded in the slow processes that underwrite experience-dependent plasticity?

Acknowledgements: KF is supported by funding for the Wellcome Centre for Human Neuroimaging (Ref: 205103/Z/16/Z), a Canada-UK Artificial Intelligence Initiative (Ref: ES/T01279X/1) and the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3).

References

1. Proietti, R., G. Pezzulo, and A. Tessari, *An active inference model of hierarchical action understanding, learning and imitation.* Physics of Life Reviews, 2023.
2. Pfister, R., et al., *Neural correlates of ideomotor effect anticipations.* Neuroscience, 2014. **259**: p. 164-71.
3. Limanowski, J., *(Dis-)Attending to the Body*, in *Philosophy and Predictive Processing*, T.K. Metzinger and W. Wiese, Editors. 2017, MIND Group: Frankfurt am Main.
4. Sperry, R.W., *Neural basis of the spontaneous optokinetic response produced by visual inversion.* J Comp Physiol Psychol, 1950. **43**(6): p. 482-9.
5. Gallese, V. and A. Goldman, *Mirror neurons and the simulation theory of mind-reading.* Trends Cogn Sci, 1998. **2**(12): p. 493-501.
6. Miall, R.C., *Connecting mirror neurons and forward models.* Neuroreport, 2003. **14**(17): p. 2135-7.
7. Kilner, J.M., K.J. Friston, and C.D. Frith, *Predictive coding: an account of the mirror neuron system.* Cogn Process, 2007. **8**(3): p. 159-66.
8. Friston, K., J. Mattout, and J. Kilner, *Action understanding and active inference.* Biol Cybern, 2011. **104**(1-2): p. 137-60.
9. Friston, K. and C. Frith, *A Duet for one.* Conscious Cogn, 2015. **36**: p. 390-405.
10. Parr, T. and K.J. Friston, *Generalised free energy and active inference.* Biol Cybern, 2019. **113**(5-6): p. 495-513.
11. Schmidhuber, J., *Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts.* Connection Science, 2006. **18**(2): p. 173-187.
12. Sun, Y., F. Gomez, and J. Schmidhuber, *Planning to be surprised: optimal Bayesian exploration in dynamic environments*, in *Proceedings of the 4th international conference on Artificial general intelligence*. 2011, Springer-Verlag: Mountain View, CA. p. 41-51.
13. Oudeyer, P.-Y. and F. Kaplan, *What is intrinsic motivation? a typology of computational approaches.* Frontiers in Neurorobotics, 2007. **1**: p. 6.
14. Barto, A., M. Mirolli, and G. Baldassarre, *Novelty or surprise?* Front Psychol, 2013. **4**: p. 907.
15. Schwartenbeck, P., et al., *Computational mechanisms of curiosity and goal-directed exploration.* Elife, 2019. **8**: p. e41703.
16. Ay, N., et al., *Predictive information and explorative behavior of autonomous robots.* European Physical Journal B, 2008. **63**(3): p. 329-339.
17. Sanchez, E.H., M. Serrurier, and M. Ortner *Learning Disentangled Representations via Mutual Information Estimation*. 2019. arXiv:1912.03915 DOI: 10.48550/arXiv.1912.03915.
18. Hafner, D., et al. *Action and Perception as Divergence Minimization*. 2020. arXiv:2009.01791.
19. Wurtz, R.H., et al., *Thalamic pathways for active vision.* Trends Cogn Sci, 2011. **15**(4): p. 177-84.
20. Sajid, N., et al., *Paradoxical lesions, plasticity and active inference.* Brain Commun, 2020. **2**(2): p. fcaa164.
21. Parr, T., N. Sajid, and K.J. Friston, *Modules or Mean-Fields?* Entropy (Basel), 2020. **22**(5): p. 552.