Commentary on: Marcel Binz, Ishita Dasgupta, Akshay K. Jagadish, Matthew Botvinick, Jane X. Wang, and Eric Schulz, "Meta-Learned Models of Cognition".

Word counts:

Abstract: 51
Main text: 835
References: 536
Entire document: 1539

Metalearning goes hand-in-hand with metacognition

Chris Fields[1] and James F. Glazebrook[2]

[1]Allen Discovery Center, Tufts University, Medford, MA 02155, USA

fieldsres@gmail.com; +33 6 44 20 68 69; https://chrisfieldsresearch.com
ORCID: 0000-0002-4812-0744

[2]Department of Mathematics and Computer Science, Eastern Illinois University, Charleston, IL 61920–3099, USA, and Adjunct Faculty (Mathematics), University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

jfglazebrook@eiu.edu; +1 217-328-4842; https://faculy.math.illinois.edu/ glazebro/
ORCID: 0000-0001-8335-221X

Abstract: Binz et al. propose a general framework for metalearning and contrast it with built-by-hand Bayesian models. We comment on some architectural assumptions of the approach, its relation to the active inference framework, its potential applicability to living systems in general, and the advantages of the latter in addressing the explanation problem.

Binz et al. craft a comprehensive outline for advancing meta-learning (MetaL) on the basis of several arguments concerning the tractability of optimal learning algorithms, manipulation of complexity, and integration into the rational aspects of cognition, all seen as basic requirements for a domain-general model of cognition. Architectural features include an inductive process from experience driven by repetitive interaction with the environment, necessitating i) an inner loop of 'base learning', and ii) an outer loop (or MetaL) process through which the system is effectively trained by the environment to ameliorate its inner loop learning algorithms. A key aspect of the model is its dependence on the relation between the typical duration of a (general, MetaL) problem-solving episode and the typical duration of a (particular, learned) solution.

While Binz et al. focus on MetaL as a practical methodology for modeling human cognition, it is also interesting to ask how MetaL as Binz et al. describe it, fits into the conceptual framework of cognition in general, and also to ask how it applies both to organisms other than humans and to artificial (or hybrid) systems operating in task environments very different from the human task environment. From a broad perspective, MetaL is one function of metacognition (e.g. Flavell, 1979; Shea and Frith, 2019;

Cox, 2005). Both MetaL and metacognition more generally engage memory and attention as they are neurophysiologically enacted by brain regions including the default mode network (Glahn et al., 2010), as reviewed for the two theories in (Wang, 2021) and (Kuchling, Fields and Levin, 2022), respectively.

When MetaL is viewed as implemented by a metaprocessor that is a proper component of a larger cognitive system, one can ask explicitly about the metaprocessor's task environment and how it relates to the larger system's task environment. MetaL operates in a task environment of learning algorithms and outcomes, or equivalently, a task environment of metaparameters and test scores. How the latter are measured is straightforward for a human modeler employing MetaL as a methodology, but is less straightforward when an explicit system-scale architecture must be specified. The question in this case becomes that of how the object-level components of a system use the feedback received from the external environment to train the metaprocessor. The answer cannot, on pain of infinite regress, be MetaL. The relative inflexibility of object-level components as "trainers" of their associated metaprocessors effectively bakes in some level of non-optimality in any multilayer system.

Binz et al. emphasize that MetaL operates on a longer timescale than object-level learning. Given a task environment that imposes selective pressures with different timescales, natural selection will drive systems toward layered architectures that exhibit MetaL (Kuchling, Fields and Levin, 2022). Indeed the need for a "learning to learn" capability has long been emphasized in the active-inference literature (e.g. Friston et al., 2016). Active inference under the free-energy principle (FEP) is in an important sense "just physics" (Friston, 2019; Ramstead et al., 2022; Friston et al., 2023); indeed the FEP itself is just a classical limit of the principle of unitarity, i.e. of conservation of information (Fields et al., 2022; Fields et al., 2023). One might expect, therefore, that MetaL as defined by Binz et al. is not just useful, but ubiquitous in physical systems with sufficient degrees of freedom. As this is at bottom a question of mathematics, testing it does not require experimental investigation.

What does call out for experimental investigation is the extent to which MetaL can be identified in systems much simpler than humans. Biochemical pathways can be trained, via reinforcement learning, to occupy different regions of their attractor landscapes (Biswas et al, 2021; 2022). Do sufficiently complex biochemical networks that operate on multiple timescales exhibit MetaL? Environmental exploration and learning are ubiquitous throughout phylogeny (Levin, 2022; 2023); is MetaL equally ubiquitous? Learning often amounts to changing the salience distribution over inputs, or in Bayesian terms, adjusting precision assignments to priors. To what extent can we describe the implementation of MetaL by organisms in terms of adjustments of sensitivity/salience landscapes – and hence attractor landscapes – on the various spaces that compose their *umwelts*?

As Binz et al. point out, in the absence of a mechanism for concrete mathematical analysis, MetaL forsakes interpretable analytic solutions and hence generates an "explanation problem" (cf. Samak et al., 2021). As in the case of deep AI systems more generally, experimental techniques from cognitive psychology may be the most productive approach to this problem for human-like systems (Taylor and Taylor, 2021). Relevant to this is an associated spectrum of ideas, including how problem solving is innately perceptual, how inference is "Bayesian satisficing" not optimization (Chater, 2018; Sanborn and Chater, 2016), the relevance of heuristics (Gigerenzer and Gaissmaier, 2011; cf. Fields and Glazebrook, 2020), and how heuristics, biases, and confabulation limit reportable self-knowledge (Fields, Glazebrook and Levin, 2024). Here again, the possibility of studying MetaL in more tractable experimental systems in which the implementing architecture can be manipulated biochemically and bioelectrically, may offer a way forward not available with either human subjects or deep neural networks.

Competing Interests: The authors have no competing interests.

Funding Statement: The authors have received no funding towards this contribution.

References:

Biswas, S., Manika, S., Hoel, E. and Levin, M. (2021) Gene regulatory networks exhibit several kinds of memory: Quantification of memory in biological and random transcriptional networks. *IScience* 24, 102131.

Biswas, S., Clawson, W. and Levin, M. (2022) Learning in transcriptional network models: Computational discovery of pathway-level memory and effective interventions. *Int. J. Molec. Sci.* 24, 285.

Chater, N. (2018). *The Mind is Flat. The remarkable shallowness of the improvising brain*. Yale University Press, New Haven and London.

Cox, M. T. (2005). Metacognition in computation : A selected research review. *Artif. Intell*. 169, 104–141.

Fields, C., Fabrocini, F., Friston, K. J., Glazebrook, J. F., Hazan, H., Levin, M. and Marcianò, A. (2023). Control flow in active inference systems, Part I: Classical and quantum formulations of active inference. *IEEE Trans. Mol. Biol. Multi-Scale Comm*. 9, 235–245.

Fields, C., Friston, K. J., Glazebrook, J. F. and Levin, M. (2022) A free energy principle for generic quantum systems. *Prog. Biophys. Mol. Biol.* 173, 36–59.

Fields, C. and Glazebrook, J. F. (2020). Do Process-1 simulations generate the epistemic feelings that drive Process-2 decision making? *Cogn. Proc.* 21, 533–553.

Fields, C. Glazebrook, J. F. and Levin, M (2024). Principled limitations on self-representations for generic physical systems. In review.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *Am. Psychol.* 34, 906.

Friston, K. J. (2019) A free energy principle for a particular physics, Preprint arxiv:1906.10184.

Friston, K. J., Da Costa, L., Sakthivadivel, D. A. R., Heins, C., Pavliotis, G. A., Ramstead, M. J. and Parr, T. (2023) Path integrals, particular kinds, and strange things. *Phys. Life Rev.* 47, 35–62.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J. and Pezzulo, G. (2016). Active inference and learning. *Neurosci. Biobehav. Rev.* 68, 862–879.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annu. Rev. Psych.*,62, 451-–482.

Glahn, D. C. et al. (2010) Genetic control over the resting brain. *Proc. Natl. Acad. Sci. USA* 10(7), 1223–1228.

Kuchling, F., Fields, C. and Levin, M. (2022). Metacognition as a consequence of competing evolutionary time scales. *Entropy* 24, 601.

Levin, M. (2022) Technological approach to mind everywhere: An experimentally-grounded framework for understanding diverse bodies and minds. *Front. Syst. Neurosci.* 16, 768201.

Levin, M. (2023) Darwin's agential materials: Evolutionary implications of multiscale competency in developmental biology. *Cell. Molec. Life Sci.* 80(6), 142.

Ramstead, M. J.,Sakthivadivel, D. A. R., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., Klein, B. and Friston, K. J. (2022) On Bayesian mechanics: A physics of and by beliefs. *Interface Focus* 13(2923), 2022.0029.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J. and Müller, K.-R. (2021) Explaining deep neural networks and beyond: A review of methods and applications. *Proc. IEEE* 109, 247–278.

Sanborn, A. N. and Chater, N. (2016). Bayesian brains without probabilities, *Trends Cogn. Sci.* 20(12), 883–893.

Shea, N. and Frith, C. D. (2019). The global workspace needs metacognition. *Trends Cogn. Sci.* 23, 560–571.

Taylor, J. E. T. and Taylor, G. W. (2021). Artificial cognition: How experimental psychology can help generate artificial intelligence. P*sychonom. Bull. Rev.* 28, 454–475.

Wang, J. X. (2021). Meta-learning in artificial and natural intelligence. *Curr. Opin. Behav. Sci.* 38, 90–95.