approach (Székely et al., 2024). In general, a combination of flexible nonlinear Bayesian models with structure learning is particularly appealing and has proven to be a valuable tool in continual learning (Achille et al., 2018; Rao et al., 2019).

References

Achille, A., Eccles, T., Matthey, L., Burgess, C. P., Watters, N., Lerchner, A., & Higgins, I. (2018). Life-long disentangled representation learning with cross-domain latent homologies. NeurIPS.

Csikor, F., Meszéna, B., & Orbán, G. (2023). Top-down perceptual inference shaping the activity of early visual cortex. *BioRxiv*. https://doi.org/10.1101/2023.11.29.569262

Éltető, N., Nemeth, D., Janacsek, K., & Dayan, P. (2022). Tracking human skill learning with a hierarchical Bayesian sequence model. *PLoS Computational Biology*, *18*(11), e1009866. https://doi.org/10.1371/journal.pcbi.1009866

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2021). Contextual inference underlies the learning of sensorimotor repertoires. *Nature*, *600*, 489–493. https://doi.org/10.1038/s41586-021-04129-3

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *PNAS*, *105*(31), 10687–10692. https://doi.org/10.1073/pnas.0802631105

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. Retrieved from www.sciencemag.org

Nagy, D. G., Török, B., & Orbán, G. (2020). Optimal forgetting: Semantic compression of episodic memories. *PLoS Computational Biology*, *16*(10), e1008367. https://doi.org/10.1371/journal.pcbi.1008367

Rao, D., Visin, F., Rusu, A. A., Teh, Y. W., Pascanu, R., & Hadsell, R. (2019). Continual unsupervised representation learning. NeurIPS.

Spens, E., & Burgess, N. (2024). A generative model of memory construction and consolidation. *Nature Human Behaviour*, *8*, 526–543. https://doi.org/10.1038/s41562-023-01799-z

Székely, A., Török, B., Kiss, M. M., Janacsek, K., Németh, D., & Orbán, G. (2024). Identifying transfer learning in the reshaping of inductive biases. *PsyArxiv*.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318. https://doi.org/10.1016/j.tics.2006.05.009

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Török, B., Nagy, D. G., Kiss, M., Janacsek, K., Németh, D., & Orbán, G. (2022). Tracking the contribution of inductive bias to individualised internal models. *PLoS Computational Biology*, *18*(6), e1010182. https://doi.org/10.1371/journal.pcbi.1010182

Meta-learning and the evolution of cognition

Walter Veit[a]* and Heather Browning[b]

[a]Department of Philosophy, University of Reading, Reading, UK and
[b]Department of Philosophy, University of Southampton, Southampton, UK
wrwveit@gmail.com
DrHeatherBrowning@gmail.com
https://walterveit.com/
https://www.heatherbrowning.net/

*Corresponding author.

Abstract

Meta-learning offers a promising framework to make sense of some parts of decision-making that have eluded satisfactory explanation. Here, we connect this research to work in animal behaviour and cognition in order to shed light on how and whether meta-learning could help us to understand the evolution of cognition.

Computational models of learning were historically largely designed by hand. But this has changed dramatically in the last decade with the rise of so-called meta-learning models that have their priors updated through feedback with the environment, thus offering a better approximation of how human cognition works due to the inclusion of flexibility and agency. Indeed, they have been able to explain some puzzling phenomena that had resisted satisfactory explanations. Yet, this wealth of research has not been unified into anything like a coherent account. The lack of such an account motivated Binz et al. to offer a synthesis of this research and framework for future research by drawing on Bayesian inference models of cognition and rational models of cognition (e.g., Anderson, 2013).

In this commentary, we do not aim to challenge their proposal, which we find very compelling. In synthesising the scattered literature and explaining meta-learning in an accessible manner, we believe the authors to be more than successful, and we share their optimism for applications of meta-learning models of cognition. Instead of criticising an aspect of their approach, we will here follow up on a question they themselves raise, but do not pursue further: "How much of it [meta-learning] is based on evolutionary or developmental processes?" (2024, p. 11). We hope to aid both the development of better meta-learning models as well as a better understanding of human learning by investigating the evolution of meta-learning from simple animals to humans. As Dennett (1995) once put it, natural selection is an acid that leaves nothing untouched, and meta-learning as we shall argue is no exception.

Binz et al. show their optimism for how meta-learning can help in understanding how cognition can develop in agents through repeated interactions with the environment, which can provide a useful model to understand human developmental processes, though they admit more research would be needed. But more interestingly perhaps – and not surprising to anyone who emphasises that development often recapitulates evolutionary processes – there is also the potential to use meta-learning models to help us understand the evolution of cognition more generally. Binz et al. urge us to consider the more complex tasks we find in natural settings for humans, but that point is worth extending towards non-human animals. While they note that meta-learning models may help us to bridge the two traditions of connectionism and Bayesian learning, an evolutionary perspective could help us to merge these traditions.

If we ask why cognition evolved – or here more specifically why creatures may have evolved meta-learning capacities – we can draw on the aforementioned puzzling tasks that meta-learning helps us to explain, such heuristic-based decision-making, as some of the authors have noted elsewhere (Binz, Gershman, Schulz, & Endres, 2022). Non-human animals, after all, also use heuristic strategies to navigate their environments. Admittedly, animal models of behaviour typically satisfy themselves with "hand-designed" algorithms of behaviour, but such models are deliberately simple to account for trade-offs between

particular considerations, for example, optimal foraging under conditions of high predator-density. Studies of animal cognition have already established that animals can solve more complex problems than was predicted (Andrews & Monsó, 2021). When Binz et al. describe the four advantages a meta-learning model has over a standard Bayesian model, two key features emerge that are highly relevant to an evolutionary account of meta-learning: Resource limitations, and the lack of prior information about the environment. When considering both of these features and the way they operate on organisms in the wild, the ecological plausibility of even an early evolution of meta-learning capacities becomes quite plausible.

Meta-learning models are able to limit the complexity of the algorithms they use, to reduce strain on resources. Under the constraints of natural selection, resource limitations play a strong role in determining the optimal strategy for organismal behaviour and/or phenotype. Out in the world, organisms will have constraints on brain size (and subsequently, memory capacity and processing power), as well as the time and energy availability for running cognitive computations. A system that provides a method for limiting the complexity of more difficult algorithms – as meta-learning does – will therefore have a strong advantage for organisms operating under normal constraints.

It is also a common feature of the environments in which animals find themselves that they will lack prior information about these environments (Veit, 2023). For any animal that lives in a variable or changing environment, or those with a complex and flexible behavioural repertoire, they cannot know in advance what they will encounter throughout their lifetimes, such as the distributions of the types of functions they will come across. A learning model that allows an organism to improve its learning over repeated encounters with and sampling of its environment will be selectively advantageous in these contexts as they can adapt to whatever circumstances in which they find themselves. Conversely, animals who evolve and develop within stable environments with a fairly fixed set of challenges may do better with pre-set learning algorithms that are optimised for this environment, to avoid the complexity and time investment required for meta-learning.

A meta-learning perspective on the evolution of animal cognition also fits with our current neuroscientific knowledge on cognitive architectures, as well as the empirical data on animal learning. For instance, Binz et al. note that many species (including humans) have been shown to improve their learning strategies over time. This empirical evidence supports the evolutionary story we have sketched here. Unfortunately, much work remains to be done in order to understand the evolution of cognition, but we hope to have successfully shown that meta-learning could offer a promising framework for enhancing such understanding, due to its inherent link to the adaptive agency of living systems.

References

Anderson, J. R. (2013). *The adaptive character of thought*. Psychology Press.

Andrews, K., & Monsó, S. (2021). Animal cognition. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2021/entries/cognition-animal/

Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2022). Heuristics from bounded meta-learned inference. *Psychological Review*, 129, 1042–1077. https://doi.org/10.1037/rev0000330

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meaning of life*. Simon and Schuster.

Veit, W. (2023). *A philosophy for the science of animal consciousness*. Routledge.

The reinforcement metalearner as a biologically plausible meta-learning framework

Tim Vriens[a] ⓘ, Mattias Horan[b] ⓘ,

Jacqueline Gottlieb[c,d]* ⓘ and Massimo Silvetti[a]

[a]Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy; [b]Sainsbury Wellcome Centre, University College London, London, UK; [c]Department of Neuroscience, Columbia University, New York, NY, USA and [d]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA
Tim.Vriens@unicampus.it,
mattias.horan.19@ucl.ac.uk,
jg2141@columbia.edu,
massimo.silvetti@istc.cnr.it
https://zuckermaninstitute.columbia.edu/jacqueline-gottlieb-phd
https://ctnlab.it/index.php/massimo-silvetti/, https://www.istc.cnr.it/en/people/massimo-silvetti

*Corresponding author.

Abstract

We argue that the type of meta-learning proposed by Binz et al. generates models with low interpretability and falsifiability that have limited usefulness for neuroscience research. An alternative approach to meta-learning based on hyperparameter optimization obviates these concerns and can generate empirically testable hypotheses of biological computations.

Binz et al. describe four different meta-learning approaches and focus on the last one – methods for learning arbitrary new tasks without the need for a priori hypotheses about brain or cognitive architectures. They show that this approach can be implemented in recurrent neural networks (RNNs) that are universal approximators (Hornik, Stinchcombe, & White, 1989), and argue that it is powerful in producing Bayesian (near-optimal) learning in an arbitrarily large set of cognitive tasks. While acknowledging the power of the proposed framework for artificial intelligence (AI), we question its usefulness in cognitive and neuroscience research. We argue that an alternative approach of hyperparameter optimization (which was first proposed by Doya, 2002, and is mentioned but not discussed by Binz et al.) is far more powerful for this role.

To be valuable for empirical research, a computational framework should generate models that are interpretable in neurocognitive terms and make predictions that can be falsified or confirmed through empirical tests. The internal computations used by the models should be analogous to those of neurocognitive systems (e.g., attention, memory, valuation, etc.; e.g., Castelvecchi, 2016), and predict activity patterns that can be empirically validated. The framework advocated by Binz et al. has neither property, and instead generates models that are