Bayes beyond the predictive distribution

Anna Székely[1,2], Gergő Orbán[1]

1, Department of Computational Sciences, HUN-REN Wigner Research Center for Physics,
Konkoly-Thege Miklós út 29-33., 1121, Budapest, Hungary

2. Department of Cognitive Science, Faculty of Natural Sciences,
Budapest University of Technology and Economics,
Műegyetem rkp. 3., H-1111 Budapest, Hungary

23 September 2024

to appear in: Behavioral and Brain Sciences
in response to: Binz M., Dasgupta, I., Jagadish, A.K., Botvinick, M., Wang J.X., &vSchulz, E. (2024) Meta-learned models of cognition. Behavioral and Brain Sciences. 2024;47:e147. doi:10.1017/S0140525X23003266

Abstract

Binz et al. argue that meta-learned models offer a new paradigm to study human cognition. Meta-learned models are proposed as alternatives to Bayesian models based on their capability to learn identical predictive distributions. In our commentary, we highlight a number of arguments that reach beyond a predictive distribution-based comparison, offering new perspectives to evaluate the advantages of these modeling paradigms.

In their review, Binz et al. propose a framework for studying the adaptive nature of the mind. They propose that recent advances in machine learning empower meta-learning paradigms to be used as a flexible and general framework for studying the computations, the representations, and even the neuronal processes underlying learning. The authors put forward a number of arguments that provide support for such a paradigm. In this commentary, we aim to reflect on these arguments in order to better identify the advantages and limits of using meta-learned models instead of Bayesian ones.

The authors pit the meta-learning paradigm against Bayesian approaches. Bayesian models provide a similarly general framework for formulating learning problems as meta-learned models, but the two paradigms differ in the principles that guide model construction. In contrast with the primarily data-driven approach of meta-learned models, Bayesian approaches formulate the computational challenge humans face when performing task(s) through the definition of likelihood and priors, which summarize our assumptions about the relevant quantities of the computational challenge and our prior beliefs about these quantities. In other words, when constructing a Bayesian model, one needs to define a generative model of the task and also the relevant quantities that shape the learning procedure, which instantly provides a set of testable hypotheses and, thus, an opportunity to better understand cognition. The authors challenge the Bayesian approach by pointing out that in complex tasks, both defining and evaluating the likelihood can be impossible, and the function classes that Bayesian models rely on can be severely constrained. The authors argue that these challenges can be circumvented by using meta-learned models instead. To support the paradigm shift, the authors cite promising new studies that explore the equiv-

alence of meta-learned models and Bayesian approaches. While these unifying views certainly contribute to a better understanding of learning, some aspects of these views deserve further consideration.

The authors argue that it is the posterior predictive distribution that a model ultimately learns, and thus, this quantity provides a platform to compare alternative approaches. The posterior predictive distribution is then used to establish the equivalence of Bayesian and meta-learned models. We would challenge this view based on two observations. First, it is important to point out that in its general form, the posterior predictive distribution is not a quantity that is invariant for a set of tasks, but it depends on the choice of the prior. This also means that the equivalence of the meta-learner and the Bayesian learner is constrained. This constraint can be illuminated by considering the contribution of the priors in Bayesian models. The effect of prior is most pronounced when data is scarce. In such cases, the equivalence is hard to establish as it is unclear what sort of prior the meta-learner model implicitly assumes. When data is abundant, however, the contribution of the prior diminishes, and in such cases, it is easier to establish the equivalence of the two model classes. Second, comparing Bayesian models and deep networks based on predictive performance alone ignores the power of having a framework that permits combining structured knowledge representations with powerful inference (Griffiths et al., 2010; Kemp et al., 2007; Kemp & Tenenbaum, 2008; Tenenbaum et al., 2006, 2011). A key benefit of Bayesian modeling is the characterization of generative models that could plausibly account for the behavioral outcomes. Creating and testing hypotheses regarding these generative models enables us to better understand the computations that underlie cognition and give rise to the behavioral outcome.

The authors refer to inductive biases that can be transparently captured by meta-learned models, some of which are not necessarily easy to capture in Bayesian models. While we agree that some forms of inductive biases are readily delivered by these meta-learned models, Bayesian models too are capable of investigating relevant inductive biases. These inductive biases might include assumptions about the function classes that learning operates on (Kemp & Tenenbaum, 2008) or assumptions about the computational complexity of the generative model (Csikor et al., 2023) both of which can be phrased through the definition of the likelihood. Such inductive biases can be explored by pitting them against alternatives and assessing the models' power to predict human learning. In summary, we argue that characterization of learning through the specification of the generative model, comprised of the prior and the likelihood, makes it possible to explore the assumptions behind the models, which assumptions may remain hidden in meta-learned models.

Finally, it's important to clarify that we agree with the authors that more flexible tools provide unique opportunities to study a broader class of phenomena. However, recent advances in Bayesian models open new opportunities in this aspect, e.g. variational autoencoders (Nagy et al., 2020; Spens & Burgess, 2024), non-parametric methods (Éltető et al., 2022; Heald et al., 2021; Török et al., 2022), or probabilistic programming (Lake et al., 2015), might leverage the need to meticulously define model architectures a priori by the experimenter and will complement the data-driven meta-learning approach proposed by the authors. In particular, the contribution of changing inductive biases to task performance in humans has been recently investigated in an implicit learning paradigm using a non-parametric Bayesian 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. N*eurIPS*.

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). 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.

Nagy, D. G., Török, B., & Orbán, G. (2020). Optimal forgetting: Semantic compression of episodic memories. *PLoS Computational Biology*, 16(10). 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*. 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). https://doi.org/10.1371/journal.pcbi.1010182