Meta-Learning: A Unifying Framework for Testing Theories of Human Learning

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Akshay Kumar Jagadish

aus Bengaluru/Indien

Tübingen

2025

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	17.09.2025
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Peter Dayan
2. Berichterstatter:	Prof. Dr. Stefano Palminteri

Do not pray for an easy life, pray to be a strong person.

– Appa

Acknowledgements

It is often said that *it takes a village to raise a child*; likewise, completing a Ph.D. would not have been possible without the support and encouragement of many mentors, friends and family. I am deeply grateful to those who instilled hope when I was dejected, lifted me when I had fallen, challenged me when I was presumptuous, and inspired me to become not only a better scientist, but also a better human being.

Tübingen: the University and the Town
I would first like to thank the University of Tübingen, and in particular the Graduate Training Center for Neuroscience, for giving me the opportunity to pursue my master's degree when I had nowhere else to turn. I am also grateful to the charming town of Tübingen for serving as the backdrop for this journey. Without its warmth and coziness, the friendships and relationships that have shaped my time here would not have taken root in the same way.

Supervisors: Eric and Marcel
Since our chance encounter at the coffee machine at the MPI, where you pitched me meta-learning, Eric, you have been incredibly inspiring and supportive. You've let me fail and succeed—on my own terms. Not all supervisors offer that kind of freedom, and I'm deeply grateful for it. You've taught me to think big, both in science and in life—something I'll carry with me always. Thank you again for giving me the opportunity to pursue my Ph.D. in your lab. While it's important to think big, it's just as crucial to be pragmatic—knowing how to break down daunting goals into actionable steps. Marcel, you epitomize this. Your consistency, dedication, and clarity of thought are second to none. Thank you for your mentorship and friendship. Together, you've carefully hand-crafted a meta-learning distribution that—I hope—will help me generalize well as a scientist at inference time.

Advisors
I would like to thank Peter and Stefano for showing me—through your actions and advice—what it means to do good science, and how to pursue it with both grace and generosity. Having already shared many intellectually stimulating conversations with you outside of any formal role, it felt completely natural to have you both transition into official members of my advisory committee. I would also like to thank Jane for helping me train my first meta-learning agent and co-advising on my first project.

Members of the CPI → HCAI lab
It is not always the case that colleagues become friends and vice versa, and I feel lucky to have had that experience. I would like to thank my contemporary, Tankred, for being part of my academic journey from master's to Ph.D. Julian, for our stimulating discussions about life and work, not to mention our many calisthenic sessions. Kristin for the fun banter during pandemic Zoom calls that continued well into the Ph.D. Mirko for our long runs and discussions about sports, food, and everything "interesting" in between. My current and former officemates—Susanne, Tobias, and Marvin—for always being open to impromptu discussions on all manner of topics. Johannes, from a student I once supervised to one of my closest friends, I'm grateful for your open, positive, and kind presence. Kozzy, for your insatiable appetite for knowledge—for reading through my proposals and dissertation drafts multiple times without hesitation. My thought partner in the past few months has been Milena—thank you for bringing me on board the journey toward automating cognitive science, and for last minute help with my figures. I also thank other lab members—Alex, Alireza, Can, Elif, Franzi, Luca, Lion, Shuchen, and Xin—for many shared experiences, from retreats and conferences to long lunches, table tennis games, and restaurant excursions.

Softhardcore
The best friendships are often the ones where you can't quite remember how they began. That was the case with this group: Erola, Sahel, Wy Ming, Simon, Berni, Stefano, Philipp, and Guillem. Each of you brought something valuable—something unique and hard to put into words—into my life, and I will always cherish that. Erola, for challenging norms and showing us that no ailment can hold back growth. Sahel, for creating an open space to raise and explore any topic under the sun. Wy Ming, for your endless patience in reading my

reports—from the start of the master's to this very dissertation. Simon, for being my personal physiotherapist and, more importantly, our cycling trip navigator. Stefano and Philipp, for your kind, loving, and generous spirit. Zeynep and Hongie, as honorary neighbors to this group, for insisting on calm evenings amid the rush. And finally, Guillem—the glue that held us together. The kind of person everyone hopes to have in their life. Thank you for keeping me grounded and for always reminding me of what really matters.

Wohngemeinschaft (WG)
After moving to Germany, I lost my sense of home for a while—until I moved into the Lemberg WG. I would like to thank Nils and Ronja for pushing my case and welcoming me in, and Kai, Erica, and Dan for later becoming part of it. The three years at Lembergstraße 11 were the most fun I have had living with friends. I thoroughly enjoyed our long chats in the bathroom, the Romania bike trip with Erica, the ridiculous challenges—from swirling frisbee throws to intense rounds of table tennis—with Nils, and the Hitchcock movie nights (and the discussions after), hikes in the Swabian Alps, and random musings about life and sports with the Dan(cer). I also thank my WG in Munich with Christin and Christina. Thank you both for your kind spirit and endless patience, especially amidst the chaos of the last few months.

Tigers
Who said tigers are solitary animals? I would like to thank this streak of tigers—Mohammad, Konstantin, Cveti, Jesse, Manos, and Sina—for chasing good times with me. From our high moments at the base of the Alps to our long chats set to great music scores, these memories are forever etched in my mind.

Longestu Bandhana
True to its meaning, this friendship has only grown stronger and deeper over the years. You remind me time and again that intense cricket games, stupid jokes, tasty meals, and strong camaraderie are all you need for a good life. Thank you, Karthik, Gagan, Azeem, Nihal, Abhi, Sohan, Suprith, and Kiran.

NLO Boys
From picking me up at the airport to being the last ones I share a drink with before boarding my flight, you've made my visits back home light and fun. Thank you, Rajath (Mama), Adi, Jagga, Sanjeev, Srini, and Sharan.

Adopted family
For people who are always shifting home bases, like me, it's crucial to have people who adopt you and make you feel part of their family. The first person who comes to mind is Yashas—a brother from another mother. Who knew that the innocuous link you sent me about MPI-IS in Tübingen would lead to this journey? Thanks for being there, my friend. Aman and Shahd—for our long weekend calls, which always left me inspired and wanting to do more and give more. I would also like to thank the Heidiris—especially Helga and Andreas—for hosting me and making me feel part of your lovely family. And special thanks to Omi, Franziska, for always welcoming us with great food, baked good(s), and a positive mood.

Family back home
Despite being an only child, when people ask if I have siblings, I always say *yes*. Thank you, Souju, Guddu, Harisha, Dacchu, and Yashu, for being the siblings I never had. I'm grateful for the blessings of my grandparents—Nagappa Thatha, Nagarathna Ajji, and Giriyamma Ajji—and of my second parents growing up, Giri Gowda uncle and Hema aunty.

Partner-in-crime
Some people just turn things for the better when they enter your life. Laura, you've been that person for me. From desert storms to thunderstorms, we've "weathered" our share of extremes—but none compare to the burden you've carried over the past months. Thank you for your belief, patience, support, and encouragement—without which this journey would have been much stormier.

Parents
How do I even begin to thank you, Amma and Appa? Nothing I say could ever do it justice. The sacrifices you've made to get me here are beyond anything I can repay. Thank you, Appa, for being my pillar of support. Your quiet belief has meant more than a thousand words ever could. Thank you, Amma, for encouraging me to pursue my goals with passion, for pushing me to explore when I was uncertain—but never too much. Thank you for instilling in me the belief that nothing is impossible if you put your heart and mind to it.

Abstract

From integrating into a foreign country to picking up a fast-paced sport, the ability to learn shapes key aspects of our everyday lives. It makes these feats seem so deceptively simple that it conceals the remarkably complex machinery that underlies them. The depth of this complexity becomes even more evident given the fact that, despite significant investments of time, effort, and money, we have yet to build machines that think, act, and learn as quickly and robustly as humans.

Over the years, cognitive scientists have developed a wide range of frameworks to understand the factors that drive human learning. While some have emphasized the interplay between goals and environments, others have focused on mechanistic constraints as key determinants of learning. Yet, none of these frameworks has succeeded in offering a complete account of human learning. What is missing is a unifying framework that simultaneously incorporates goals, environmental structure, and cognitive constraints—crucially, one that enables models to *learn how to learn*, rather than being hand-crafted for individual tasks.

In this thesis, I argue that meta-learning provides such a framework. By training systems, typically neural networks, to adapt to tasks drawn from structured environments under constraints, meta-learning offers a computational instantiation of learning that is flexible, efficient, bounded, and adaptive.

I begin this thesis by illustrating how meta-learning (1) supports the construction of sample-efficient learning algorithms that implement Bayes-optimal policies; (2) allows for easy manipulation of computational complexity of so derived learning algorithms; (3) enables precise control over adaptive environments; and (4) facilitates the integration of neuroscientific insights into model architecture. In doing so, the resulting models combine the strengths of two classic computational modeling approaches: Bayesian and connectionist models. Subsequently, I showcase the unique capabilities of meta-learning through four different studies.

First, I use meta-learning to perform a rational analysis of optimism bias—the tendency to learn more from better-than-expected outcomes than from worse-than-expected ones. Using idealized learning algorithms derived via meta-learning, I demonstrate that this tendency can emerge as a rational strategy, shaped by optimal adaptation to the diverse environments previously observed or interacted with.

Second, I provide a resource-rational account of human compositional reinforcement learning. I begin by showing that neural networks trained via meta-learning can perform near-optimal compositional inference in structured bandit tasks. Then, I demonstrate that a resource-limited version of the same model is necessary to capture the deviation from optimal inference observed in human behavior, revealing that the extent of human compositional reasoning is limited by their cognitive capacity.

Third, I derive models of ecological rationality using meta-learning to explain human category learning. I show that large language models (LLMs) can generate ecologically valid classification tasks at scale and that models adapted to the statistics of these generated tasks capture key aspects of human categorization, including learning difficulties, learning strategies, and patterns of generalization.

Fourth, I introduce ecologically rational analysis, a new computational framework that unifies the normative foundations of rational analysis with the ecological grounding of heuristic models. I show that models derived using this framework—by leveraging meta-learning to distill ecological priors into neural networks—account for a substantial portion of human behavior across three core cognitive domains: function learning, category learning, and decision making.

Overall, my findings suggest that meta-learning offers a domain-general framework for testing theories of human learning. In doing so, it provides a unifying perspective that bridges Bayes and connectionism, optimality and boundedness, and structure and adaptation. Looking ahead, this framework brings us closer to a general theory of human learning—one that explains how we adapt and thrive in the real world, whether by navigating an unfamiliar culture or mastering a demanding sport.

Kurzfassung

Von der Integration in ein fremdes Land bis zum Erlernen einer anspruchsvollen Sportart – die Fähigkeit zu lernen prägt zentrale Aspekte unseres Alltags. Sie lässt diese Leistungen so mühelos erscheinen, dass sie die zugrunde liegende, bemerkenswert komplexe kognitive Architektur verschleiert. Wie tiefgreifend diese Komplexität ist, zeigt sich auch daran, dass wir trotz erheblicher Investitionen an Zeit, Aufwand und Ressourcen noch keine Maschinen entwickelt haben, die ebenso schnell und robust denken, handeln und lernen wie der Mensch.

Im Laufe der Jahre haben Kognitionswissenschaftler*innen zahlreiche theoretische Rahmen entworfen, um die Prinzipien menschlichen Lernens zu verstehen. Einige betonen das Zusammenspiel von Zielen und Umwelt, andere konzentrieren sich auf kognitive Begrenzungen als treibende Faktoren. Dennoch ist es bislang keinem Ansatz gelungen, menschliches Lernen umfassend zu erklären. Was fehlt, ist ein integrierender Rahmen, der Ziele, Umweltstruktur und kognitive Einschränkungen zugleich berücksichtigt – und es Modellen ermöglicht, to learn how to learn, anstatt auf einzelne Aufgaben zugeschnitten zu sein.

In dieser Arbeit zeige ich auf, dass meta-learning einen solchen Rahmen bietet. Durch das Trainieren von Systemen – typischerweise neuronalen Netzwerken – auf Aufgaben aus strukturierten Umgebungen unter Ressourcenbeschränkungen entsteht eine rechnerische Instanziierung von Lernen, die flexibel, stichprobeneffizient, ressourcenschonend und adaptiv ist.

Ich beginne diese Arbeit mit einer Darstellung, wie meta-learning (1) die Entwicklung stichprobeneffizienter Lernalgorithmen unterstützt, welche Bayes-optimale Strategien implementieren; (2) gezielte Steuerung der computational complexity dieser Algorithmen erlaubt; (3) präzise Anpassungen an adaptive Umgebungen ermöglicht; und (4) die Integration neurowissenschaftlicher Erkenntnisse in die Modellarchitektur erleichtert. Auf diese Weise kombinieren die resultierenden Modelle die Stärken zweier klassischer Ansätze: Bayesian und connectionist models. Anschließend illustriere ich die besonderen Möglichkeiten von meta-learning anhand von vier Studien.

Zunächst nutze ich meta-learning, um eine rationale Analyse des optimism bias durchzuführen – der Tendenz, aus besser als erwarteten Ergebnissen stärker zu lernen als aus schlechteren. Anhand idealisierter, meta-gelernter Lernalgorithmen zeige ich, dass dieses Phänomen als rationale Strategie verstanden werden kann, die aus optimaler Anpassung an zuvor erlebte Umwelten resultiert.

Zweitens präsentiere ich eine resource-rational Erklärung für kompositionelles reinforcement learning beim Menschen. Ich demonstriere zunächst, dass neuronale Netzwerke, die mittels meta-learning trainiert wurden, nahezu optimale kompositionelle Inferenz in strukturierten Bandit-Aufgaben leisten. Anschließend zeige ich, dass nur eine ressourcenbegrenzte Variante desselben Modells die Abweichungen vom optimalen Verhalten beim Menschen erklären kann – und somit die kognitiven Kapazitätsgrenzen menschlicher Generalisierung aufzeigt.

Drittens leite ich mithilfe von meta-learning Modelle von ecological rationality ab, um menschliches Kategorisierungslernen zu erklären. Large language models (LLMs) generieren dafür in großem Maßstab ökologisch valide Klassifikationsaufgaben. Modelle, die auf die Statistik dieser Aufgaben abgestimmt sind, erfassen zentrale Merkmale menschlicher Kategorisierung – einschließlich Lernschwierigkeiten, Strategien und Generalisierungsmuster. Viertens stelle ich die ecologically rational analysis vor – einen neuen rechnerischen Rahmen, der die normativen Grundlagen der rational analysis mit der ökologischen Fundierung heuristischer Modelle verbindet. Die durch meta-learning herausgearbeiteten ecological priors führen zu Modellen, die einen erheblichen Teil menschlichen Verhaltens in drei zentralen kognitiven Domänen erklären: Funktionslernen, Kategorisierungslernen und Entscheidungsfindung.

Insgesamt zeigen meine Ergebnisse, dass meta-learning einen domänenübergreifenden Rahmen zur Überprüfung theoretischer Modelle des menschlichen Lernens bereitstellt. Es eröffnet eine integrierende Perspektive, die Brücken schlägt zwischen Bayes und connectionism, zwischen optimality und boundedness sowie

zwischen Struktur und Anpassungsfähigkeit. Dieser Ansatz bringt uns einer allgemeinen Theorie des menschlichen Lernens näher – einer Theorie, die erklärt, wie Menschen sich erfolgreich an die reale Welt anpassen, sei es beim Zurechtfinden in fremden Kulturen oder beim Erlernen komplexer Fertigkeiten.

Contents

What is the premise?

Background 1

1.1 Human learning is a remarkable feat

Learning is one of nature's quiet miracles. Without explicit instruction or careful planning, we improve simply by doing: refining perception, tuning motor actions, and constructing increasingly abstract representations of the world. With each repetition, we become more precise, more fluent, more expert. This capacity underlies feats as diverse as adapting to life in a new country, picking up a social dance like Bachata, or mastering a fast-paced team sport like Ultimate Frisbee. From the mundane to the extraordinary, learning is the thread that stitches experience into skill.

And yet, this seemingly effortless process conceals astonishing complexity. Human learners routinely generalize from sparse data, transfer knowledge across domains, and acquire structured representations that scaffold future learning. Despite decades of work in artificial intelligence, replicating this flexibility and efficiency remains an elusive goal [1]. The challenge is not just to learn, but to learn like us, with speed, grace, and generality.

What is remarkable is not just that humans learn, but how they learn. Human learning is extraordinarily sample-efficient, energy-efficient, and robustly self-supervised. With a single exposure, a child can acquire a new word [2]; with minimal energy, the brain performs inferences that challenge even today's most advanced systems [3]. We learn without explicit instruction [4], and we routinely generalize to unfamiliar problems with little more than structural insight [5]. These abilities reveal not just the power of human cognition, but the inadequacies of current artificial models that aspire to match it.

Over the years, cognitive scientists have proposed a wide range of frameworks to explain how humans learn. Some focus on the rational goals of learning, others on the mechanistic constraints of cognition, and still others on the structure of the environment in which learning occurs. Despite the elegance of these frameworks–ranging from rational analysis [6] to resource-rational [7] and ecologically-rational models [8]–none offers a complete account of how learning emerges from experience under realistic constraints. What is missing is a unified account that simultaneously incorporates goals, environmental properties, and cognitive limitations. Crucially, we need a framework that enables models to learn how to learn, rather than being hand-designed for each task.

In this thesis, I argue that **meta-learning** [12] provides such a framework. By training systems to adapt to tasks drawn from structured environments under constraints, meta-learning offers a computational instantiation of learning that is flexible, efficient, bounded, and adaptive. I show how this framework can be used to test and extend theories of human learning, and in doing so, suggest a unifying perspective that bridges Bayesianism and connectionism, optimality and boundedness, and structure and adaptation.

Def.: Meta-learning

The term meta-learning comprises two components: *learning*, which denotes the improvement of performance through experience [9], and the prefix *meta*, indicating a higher-order or self-referential level. Combined, meta-learning refers to the process of *learning how to learn* [10–12]

1.2 The setting

Imagine a participant arriving at a psychology lab to take part in an experiment. They are first informed about the task they will perform that day, say, category learning. Next, they receive instructions on how the experiment works: stimuli will be presented sequentially, and their goal is to learn to *correctly* categorize each one into one of two categories based on feedback received after each observation. Once these formalities are completed, the participant proceeds with the actual experiment.

Previous studies have shown that people are exceptionally good at category learning [13]. In a matter of a few trials[1], they can learn to correctly assign stimuli to their corresponding categories, using past observations and feedback. Remarkably, they can also categorize unseen stimuli into categories that best align with the true underlying structure.

1: The rate and extent of human learning depends on the precise category structure—the underlying rule or criterion used to assign stimuli to categories—used in the experiment [14].

Such sample-efficient learning and generalization have been observed in cognitive domains beyond category learning, including function learning [15], decision making [16], and reinforcement learning [17]. What drives learners to generalize so efficiently? Specifically, how does prior real-life experience shape how participants learn in these experimental settings? How do the task objective and implicit constraints, ranging from memory to attention, determine the representations they form and how they use them to make decisions?

To answer these questions, cognitive scientists have increasingly relied on cognitive modeling [18, 19]. By formulating cognitive theories and building computational models to test them, cognitive modeling sheds light on the processes and mechanisms underlying human behavior. In this thesis, I propose that meta-learning[2] offers a computational framework for exploring the factors that shape human learning, particularly those rooted in normative theories, ecological adaptation, and cognitive constraints.

2: I use meta-learning *only* to model learning on the time scale of psychology experiments, as illustrated in the example. While previous work has argued that meta-learning provides a framework for studying learning across multiple nested time scales—from evolutionary processes [20] to developmental trajectories [21]—examining learning over such extended time scales lies beyond the scope of this thesis.

1.3 Meta-learning as a radical new framework for testing theories of human learning

The framework of meta-learning turns traditional cognitive modeling approach on its head. Instead of hand-designing learning algorithms like traditional approaches, meta-learning allows us to learn the learning algorithm itself through repeated interactions with tasks sampled from an environment. In doing so, meta-learning offers a radically different approach to testing computational theories of human learning. Computational models derived using meta-learning have been shown to capture human learning across multiple domains, including decision making [22], probabilistic learning [23], and structure learning [24].

There are two predominant ways to construct models using meta-learning: optimization-based meta-learning and memory-based meta-learning [3]. Next, I will describe both these methods, contrast the approach they take, and lay the foundations for building computational models of human learning using meta-learning.

These two methods assume the model uses a neural network core. They are alternative approaches to meta-learning, for example, hierarchical Bayesian models [25, 26], but these methods are beyond the scope of this thesis.

3: Sometimes, also referred to as model-based meta-learning[27]. For an exhaustive review of meta-learning, see [28], [29], and [30].

1.3.1 Optimization-based meta-learning

Model-agnostic meta-learning (MAML), introduced by Finn, Abbeel, and Levine, was the first instantiation of optimization-based meta-learning. Its goal is to learn initial weights for a neural network such that, with only a few gradient-based updates, the model achieves near-optimal performance on new tasks drawn from a task distribution. MAML applies broadly across architectures—from multilayer perceptrons to convolutional networks—making it agnostic to network structure. Its only requirement is that model parameters be differentiable for gradient-based optimization. Neural networks trained with MAML have been shown to generalize efficiently across diverse domains, including regression, classification, and reinforcement learning [31], making it an attractive framework to build models of human learning.

To formalize this, let $\mathcal{T}_i$ denote a batch of tasks sampled from the task distribution $p(\mathcal{T})$ and let f_θ be a neural network model with parameters θ. Each task $\mathcal{T}_i$ is associated with a loss function $\mathcal{L}_{\mathcal{T}_i}$. To perform well on any newly encountered task, the parameters of the network are updated using gradient descent [32]. A single update step from the initial parameter θ to the adapted parameter ϕ is given by the following:

$$\phi = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}\left(f_\theta\right) \tag{1.1}$$

where α is the learning rate, and, in practice, ϕ denotes the updated parameters of the neural network after one or more gradient descent steps, computed over tasks sampled from $p(\mathcal{T})$.

The goal of MAML is to learn the initial parameters θ for a neural network such that, after a few gradient-based updates, the model achieves near-optimal performance on tasks sampled from $p(\mathcal{T})$. Formally, this involves finding parameters θ that minimize the loss of the adapted model $f(\phi)$. This objective can be expressed as minimizing the following meta-objective:

$$\min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_\phi\right) = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)}\right) \tag{1.2}$$

Crucially, the meta-optimization in Equation 1.2 is performed over the initial model parameters θ, while the loss within the meta-objective is computed using the adapted model parameters ϕ.

At inference time, the network parameters are initialized with those that minimized the meta-objective in Equation 1.2. Given a small dataset, consisting of input–target pairs, the model updates its parameters using gradient descent, following the procedure in Equation 1.1 for the same number of update steps used during training. After that, the network can efficiently learn the underlying input–target mapping and achieve near-optimal performance on the held-out dataset.

Taking the optimization-based approach for meta-learning comes with several key advantages. First, meta-learned models derived using this approach are extremely flexible, since it is agnostic to the model architecture and task domains, making it broadly applicable [30]. Second, they learn the knowledge implicitly within the network parameters that is easily transferable to related domains [31]. Third, they use gradient-based

The original MAML approach, as described above, was limited to learning only the initial weights of the neural network. Several extensions have since been proposed. Some variants of MAML learn not only the initial parameters, but also the learning rate and the direction of update [33]. PLATIPUS [34] introduces a Bayesian formulation by modeling a distribution over initializations, allowing task-specific uncertainty estimation during meta-learning. Other variants, such as First-Order MAML (FOMAML), simplify training by approximating second-order gradients with first-order updates [35]. Collectively, these methods are often referred to as MAML model "zoo".

updates for learning instead of black-box optimization, which supports formal analysis of learning trajectories [36–38]. Fourth, as a framework for building computational models of human behavior, they provide the means to bridge the computational level with the algorithmic level of analysis [26]. Specifically, Grant and colleagues formalized a connection between MAML and hierarchical Bayesian inference by identifying the corresponding prior, likelihood, and parameter estimation procedures used during meta-learning.

Building on these advantages, recent studies have used optimization-based meta-learning, specifically MAML, to build models that can explain qualitative aspects of human cognition. McCoy et al., for example, used MAML to learn initial weights that allows neural networks to quickly adapt to new languages. They found that the learned initial weights captured factors that are shared by the distribution of languages seen during training, effectively serving as universal inductive biases for the acquisition of language (e.g., syllable structure) [39]. Similarly, Dubey et al., extended MAML to account for the task context during meta-learning and showed that the resulting models captured "the context-sensitivity of human behavior in a simple but well-studied cognitive control task" [40].

Despite its strengths, cognitive modeling using optimization-based meta-learning comes with certain limitations. First, it restricts learning to gradient-based updates, narrowing the space of possible learning trajectories [41, 42]. Second, these approaches typically require computing second-order gradients (or their approximations), which can be computationally expensive and challenging to scale [35]. Third, it assumes that the tasks used for meta-learning share similar loss landscapes, which can lead to poor performance when this assumption is violated [30, 42]. Fourth, it does not support explicit control over computational complexity during inference (see Section 2.1 in Publications).

1.3.2 Memory-based meta-learning

Originally developed by Hochreiter, Younger, and Conwell, Memory-based meta-learning replaces gradient-based updates used in MAML with learning through internal activation dynamics at inference. The learning algorithm is typically implemented by a general-purpose function approximator, which can be any sequence model, from classic recurrent neural networks to modern architectures such as transformers [44] and structured state-space models [45]. Meta-learned models derived using MBML have also been shown to learn sample efficiently in multiple task settings, including regression [43], classification [46], and reinforcement learning [47].

Formally, turning a general function approximator into a meta-learning algorithm with MBML involves two stages: in-context learning[4] and in-weight learning. During in-context learning, a neural network with parameters θ interacts with a task sampled from a task distribution $p\left(\mathbf{x}_{1:T}, y_{1:T}\right)$. For each task, the network attempts to predict the target y_{t+1} for a given input $\mathbf{x}_{t+1}$, conditioned on all previous input-target pairs $(\mathbf{x}_{:t}, y_{:t})$. Importantly, during this phase, the network's parameters remain fixed across time steps. In the in-weight learning stage, the

4: In-context learning is a recently coined term used to describe a network's ability to rapidly learn from observations provided in its context while keeping its parameters frozen [48]. Although originally introduced to explain this ability in large language models (LLMs), meta-learned models derived using MBML exhibit the same capacity. The key difference is that MBML explicitly trains neural networks to perform in-context learning, whereas in LLMs, it emerges as a byproduct of scale of the training data, neural network size, and its internal dynamics.

network parameters are updated using gradient descent to maximize the following objective:

$$\arg\max_{\theta} \mathbb{E}_{p(\mathbf{x}_{1:T}, y_{1:T})} \left[\sum_{t=1}^{T} \log p_\theta \left(y_{t+1} \mid \mathbf{x}_{t+1}, \mathbf{x}_{:t}, y_{:t} \right) \right] \tag{1.3}$$

where p_θ denotes the output probabilities produced by the neural network. In practice, meta-learned models are trained on a sample-based estimate of Equation 1.3. During the training phase, the two stages are alternated until a convergence criterion is met. At inference, the neural network functions as an in-context learner: it predicts the target of a new input based on preceding input-target pairs, using only its internal activation dynamics for learning while keeping its parameters fixed.

Given its complementary approach to meta-learning, MBML is particularly well suited to address several limitations of the optimization-based approach discussed previously. First, learning implemented via internal network dynamics allows for richer learning rules than those offered by gradient descent [49]. Second, since inference-time learning occurs through feedforward passes rather than gradient-based updates, MBML is both more sample- and time-efficient. Moreover, it relies on standard first-order gradient descent for training, which is highly optimized for scaling [50]. Third, meta-learned models derived using MBML have been shown to implement Bayes-optimal learning algorithms[5] during in-context learning [51], enabling a direct link between meta-learning and rational analysis of cognition [52]. Fourth, MBML provides a principled way to control the complexity of the in-context learning algorithm, making it useful for testing resource-rational theories of human learning [22, 53]. I will further elaborate on these advantages in Section 1.5 and Section 1.6 in Background, and provide a more detailed discussion in Section 2.1 in Publications.

5: For a more accessible proof, see Box 1 in Section 2.1 in Publications.

Taken together, memory-based meta-learning offers several key advantages that make it an attractive framework for building computational models of human learning. The remainder of this thesis is dedicated to exploring this possibility[6].

6: Going forward, I focus exclusively on memory-based meta-learning for building cognitive models, henceforth referred to simply as meta-learning.

1.4 Meta-learning bridges Bayesian and connectionist models of human learning

The debate at the heart of this section can also be viewed through related perspectives—such as the "neats versus scruffies debate" [54], the contrast between symbolic and sub-symbolic approaches to cognition [55], or Rationalism versus Empiricism discussion [56], to name a few.

A long-standing debate in cognitive science has revolved around whether human cognition emerges from a complex interplay of simpler processes, as assumed by Emergentists, or is better viewed as a structured, goal-directed process that constructs knowledge via inference over abstract representations, as posited in constructivist theories. These philosophical perspectives map closely onto two dominant computational frameworks: Bayesian models and connectionist models. Bayesian models, typically symbolic in nature, adopt a constructivist perspective, positing that cognition involves inference over structured hypothesis and prior knowledge, treating it as a process of rational belief updating [57]. In contrast, connectionist models implement a sub-symbolic emergentist view. They let structure emerge from a dynamic interaction of simple processing

[57]: Griffiths et al. (2010), 'Probabilistic models of cognition: Exploring representations and inductive biases'

[58]: McClelland et al. (2010), 'Letting structure emerge: connectionist and dynamical systems approaches to cognition'

units without requiring explicit symbolic representations or priors [58]. These two modeling traditions thus reflect the contrasting views on how learning and representation are acquired and structured in the mind.

Each approach brings powerful modeling capabilities rooted in its foundational philosophy and has made significant contributions to our understanding of human cognition, especially in the domain of learning. For instance, Bayesian models are especially effective for modeling how people learn and generalize from limited data [59, 60]. They allow for precise modeling of how prior knowledge shapes generalization, how learners update beliefs in response to new evidence, and how inductive biases support sample-efficient learning. Given their normative ground, they have been instrumental in formalizing rational accounts of learning across domains like function learning [61], causal inference [62], and categorization [63]. Connectionist models, in turn, excel at explaining how cognitive abilities can arise from simple processing units through experience-driven adaptation [64–66]. Rather than imposing structure a priori, they allow representational and behavioral regularities to emerge from data, enabling insights into the developmental trajectory of learning [67]. These models are especially useful for capturing how cognitive processes unfold over time, simulating effects of graded capacity constraints, and providing mechanistic links to neural computation [64]. Through learned distributed representations, they have shed light on phenomena such as attentional shifts [68], the formation of semantic categories [69], and the acquisition of language [70].

To illustrate the contrast between the two approaches, let us view human category learning through their respective lens. From the Bayesian perspective, categorization can be modeled as probabilistic inference over structured hypotheses about category membership. For instance, the Rational Model of Categorization (RMC) [6] assumes that learners possess a priori knowledge of category structures, represented as clusters of feature vectors, and use Bayes rule to update their beliefs about category assignment based on observed items. RMC captures how people integrate limited evidence with structural assumptions to infer category boundaries and generalize to new items. In contrast, connectionist models like ALCOVE [71] offer a complementary account in which category learning emerges from incremental, error-driven adaptation. The model does not assume any explicit structure; rather, it learns to prioritize discriminative features and builds category representations based on repeated exposure to stimuli and feedback. As a result, categorization behavior emerges from a gradual process of tuning internal parameters, offering a mechanistic account of how structure can be learned rather than assumed.

While adopting contrasting approaches has led these frameworks to develop unique strengths, their deep commitment to them introduces an idiosyncratic set of weaknesses. For example, Bayesian models learn efficiently from limited data and generalize well to novel structures. Rooted in formal notions of rationality, they are well-suited for constructing (or approximating) rational models of human learning. However, doing so requires making apriori assumption about the structure of prior knowledge and how it updates it beliefs based on incoming information. While this offers flexibility, it also places a substantial burden on the modeler [58, 72]. Furthermore, Bayesian inference becomes computa-

tionally intractable very easily as task complexity increases, limiting the feasibility of these models in real-world settings[7]. Their reliance on discrete symbolic representations also makes it difficult to relate them to mechanisms used by the brain [74]. In contrast, connectionist models scale well to complex tasks, learn inductive biases directly from data, rely on distributed vector representations, and come with productive analogy to neural computation. However, they tend to be highly sample-inefficient, struggle with systematic generalization, suffer from catastrophic forgetting, and are often difficult to interpret [57]. Crucially, it is difficult to guarantee their convergence to optimal policies, especially is more complex settings, making them ill-suited for building rational models of cognition [75].

7: However, there currently exists a strong suite of approximate inference methods for modeling human reasoning, which have achieved considerable success—even in settings that appear computationally intractable [73].

Taken together, these limitations point to the need for a modeling framework that combines the sample-efficient reasoning and normative foundations of Bayesian models with the scalability and automaticity of connectionist systems. Meta-learning provides a promising path toward such unification. Computational models built using meta-learning learn sample-efficiently, acquire inductive biases directly from data, implement (approximately) Bayes-optimal policies, scale well to naturalistic environments, and offer compelling analogies to prefrontal cortical function [80].

Another cognitive modeling approach that has also been successful in unifying the probabilistic and connectionist models is neuro-symbolic modeling [76–78]. Hybrid neuro-symbolic models have been shown to combine symbolic, probabilistic, and neural learning methods to acquire explicit representations of domain knowledge and conceptual structure that are interpretable to humans [79].

In Section 2.1 – Meta-learned models of cognition, I discuss in detail the advantages meta-learning offers over traditional Bayesian inference methods, while preserving the representational flexibility and scalability of the connectionist approach. I also elaborate on how meta-learning supports the acquisition of structured hypotheses and priors from experience, thereby helping to build a bridge between emergentist and constructivist accounts of cognition. With regard to empirical work, in Section 2.3 – A resource-rational account of zero-shot compositional inference in a reinforcement learning setting, I will show that compositional reasoning, historically studied using Bayesian models with handcrafted symbolic priors, can be meta-learned from repeated interaction with environments that demand compositional inference for optimal performance. Later, in Section 2.5 – Meta-learning ecological priors from large language models captures human learning and decision making, I will show how symbolic priors from Bayesian models (e.g., rational model of function learning [61] and rational model of categorization [52]) can be distilled into neural networks using meta-learning, demonstrating how symbolic priors can be distilled in sub-symbolic models (see Section 3.1 in Outlook for extended discussion).

1.5 Meta-learning for rational analysis

Rational analysis has been one of the most influential frameworks for developing rational models of human behavior [52]. The framework posits that human behavior can be explained by considering "what would be optimal behavior given the structure of the environment and the goals of the human"[52]. Deriving a rational model involves specifying a precise goal for the agent, the formal model of the environment with which it interacts, and assumptions about computational limitations. The optimal behavioral function is then derived using optimization methods under

[52]: Anderson (1990), *The adaptive character of thought*

these specifications and compared with empirical behavior of humans, with the steps repeated until a convergence criterion is met; refer to Box 1.5 for detailed steps (adapted from [81]).

Recipe prescribed for constructing models of rational analysis

1. **Goals**: Precisely specify the goals of the cognitive system.
2. **Environment**: Develop a formal model of the environment to which the system is adapted.
3. **Computational limitations**: Make minimal assumptions about computational limitations.
4. **Optimization**: Derive the optimal behavioral function given steps 1 through 3.
5. **Data**: Examine the empirical literature to see if the predictions of the behavioral function are confirmed.
6. **Iteration**: If the predictions are off, iterate.

Rational analysis has been successfully applied to various cognitive domains, such as memory, categorization, inductive reasoning, language processing, and information search, demonstrating phenomena such as forgetting, category formation, reasoning biases, and linguistic patterns as adaptive responses to environmental statistics and task-specific objectives [6, 81–84]. To illustrate this concretely, consider human category learning once again. Rational analysis frames categorization as an optimal inference process, where categories are formed by integrating prior knowledge about the category structure and input observations following the Bayes rule to maximize predictive accuracy. Specifically, the rational model of categorization assumes that categories reflect environmental regularities by partitioning incoming observations (or items) into disjoint groups, each characterized by independent features. This assumption[8] simplifies the inference process, allowing optimal probabilistic estimation of the unseen feature, which in this case is the category to which an item belongs. Empirical studies have shown that human categorization is closely aligned with these optimal Bayesian predictions, supporting the adaptive nature of categorization behavior [6].

8: It was inspired by the organization of living objects into species with distinct traits [6]

Despite its elegance, the framework makes several simplifying assumptions that limit its applicability, especially when it is required to explain behavior in the wild. Most notably, rational analysis presupposes access to a complete and formal model of the environment. However, it is often infeasible to define such a model in real world settings [81], Environmental features tend to interact in complex, non-linear ways; cues may be redundant, only partially observable, or dynamically available; and the relationships between cues and outcomes are frequently stochastic. Furthermore, deriving an optimal behavioral function in such realistic settings becomes computationally intractable fairly quickly [85].

Consequently, models built using rational analysis have also frequently resorted to substantial simplifications, such as assuming stable isolated features with clear adaptive value, to maintain tractability. While effective in controlled laboratory tasks, these simplifications have limited the applicability of rational analysis to ecologically valid settings [6, 86]. This discrepancy has led to skepticism about the feasibility of using

traditional rational analysis "to build plausible models of human cognitive processes" [87]. Together, these challenges underscore the need for a new methodology to implement rational analysis. Ideally, such a modeling approach allows us to derive optimal response functions adapted directly to the properties of real-world environments rather than its simplified representations. Moreover, while doing so, it must remain computationally feasible.

Meta-learning emerges as a promising modeling approach that satisfies several key requirements: (1) it offers guarantees for implementing approximately Bayes-optimal solutions [51]; (2) owing to its neural network-based architecture, meta-learning is inherently scalable; and (3) it adapts to the statistical properties of the environments used during training, making it well-suited for richer, more complex settings while reducing the need for extensive hand-design. I elaborate on these arguments in Section 2.1 – Meta-learned models of cognition in Publications. Specifically, I first illustrate how meta-learning provides a powerful tool for constructing Bayes-optimal learning algorithms, while circumventing issues of computational intractability (and neural implausibility) typically faced by Bayesian models. I then draw connections between meta-learned cognitive models and rational analysis, and reinterpret prior work using meta-learning to build models of human cognition in light of this perspective. Later, in Section 2.2 – In-context learning agents are asymmetric belief updaters, I concretely demonstrate how meta-learning can be leveraged for rational analysis of optimistic belief updating observed during human reinforcement learning. In particular, I show that meta-learning can explain the emergence of optimism bias[9]. Taken together, these works establish meta-learning as a promising framework for the rational analysis of cognition—especially learning.

9: A tendency to weigh positive outcomes more heavily than negative ones, interpreted here as a rational response to the structure of real-world problems [88, 89]

1.6 Meta-learning for resource-rational analysis

While rational analysis considers only functional constraints, resource-rational analysis recognizes the importance of structural constraints in explaining human behavior [7]. It proposes that the behavior of learners is optimized not only for their environment but also for their inherent limitations, such as working memory capacity, attention spans, processing time, and energy resource. This shift in emphasis leads resource-rational analysis [10] to provide a more realistic account of cognition by explaining why people often rely on heuristics or simplified strategies, not as failures of rationality but as optimal adaptation to their cognitive constraints. Moreover, it provides a principled framework for explaining individual differences in behavior, since variations in cognitive resources can directly shape the strategies employed. Crucially, one of its key contributions lies in linking multiple levels of explanation: connecting abstract computational goals with algorithmic mechanisms capable of achieving those goals under realistic cognitive limitations [93].

[7]: Lieder et al. (2020), 'Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources'

10: Resource-rational analysis has its roots in the field of "bounded optimality," which aimed to develop optimal programs for agents that interact with their environment in real time while operating under performance-constrained hardware [90, 91]. This framework was later adopted by psychologists to theorize how people *should* think and reason in order to make the best use of their limited cognitive resources [7, 92, 93].

Building a resource-rational model begins by formalizing the cognitive function to be optimized and identifying algorithmic strategies that approximate this function under resource constraints. From among these,

the strategy that best balances computational cost and accuracy is chosen, often by maximizing a rational objective such as expected utility under constraints. The predictions from the winning model are finally evaluated against human behavior, and the process iteratively refines assumptions about the task, computational architecture, or resource bounds until the model captures empirical data sufficiently well; see Box 10 for an overview of the modeling process (adapted from [93]).

Recipe prescribed for constructing models of rational analysis

1. **Function**: Formalize the problem that the cognitive mechanism solves and characterize the optimal solution.
2. **Model of mental computation**: Identify a family of algorithms that approximate the optimal solution and their computational costs.
3. **Optimal resource allocation**: Find the algorithm in this class that optimally trades off approximation accuracy against time and other resources
4. **Evaluate**: Compare the predictions of the model against human behavior.
5. **Refine**: Revisit steps 1 through 3 if predictions are still off, or alternatively, proceed to a level below by modeling how the basic operations might be approximated or considering additional resource constraints.
6. **Iterate**: If the predictions are still off, iterate.

Resource-rational analysis has been productive at explaining phenomena such as chunking in memory [94, 95], hierarchical action planning [96], satisficing [97], sparse attention allocation [98], and perceptual filtering through latent inhibition [99], showing how these strategies emerge as bounded-optimal behavior balancing utility with cognitive cost [100–104]. Reconsidering the rational model of categorization (RMC) from a resource-rational perspective (for illustration). While the original formulation proposes an optimal Bayesian categorization strategy, practical constraints required Anderson to use approximation strategies. In particular, RMC assigns items to categories using an approximation strategy based on a greedy maximization algorithm. But this is just one possible approximation, [93] highlighted that RMC could be implemented using other approximation strategies, such as Markov chain Monte Carlo or particle filtering [105, 106], which are capable of approximating the average over all clusters by sampling across possible category assignments. Importantly, when these sampling algorithms are constrained by limited computational resources, such as a small number of samples, they can yield predictions that systematically deviate from ideal Bayesian predictions, producing human-like biases such as order effects. In doing so, Griffiths, Lieder, and Goodman demonstrated how resource-rational analysis effectively bridges normative predictions with cognitive limitations, enabling the construction of a "more realistic model of psychological processes and mental representations" that goes beyond mere behavior prediction [93].

As they provide a flexible means of incorporating cognitive constraints (as demonstrated above), Bayesian models have become a leading framework

for implementing resource-rational analysis. For instance, methods, such as Markov chain Monte Carlo and particle filtering, can be readily adapted to derive resource-rational algorithms [107]. However, a key limitation of these methods is that they mainly manipulate *computational complexity*, that is, the time, space, or effort required to execute an algorithm, while largely overlooking *algorithmic complexity*. Incorporating constraints on algorithmic complexity, which concerns the number of bits needed to represent the algorithm itself, poses a significant challenge for Bayesian inference. This difficulty arises because such measures, like Kolmogorov complexity, are generally non-computable to the best of my knowledge with most standard probabilistic inference frameworks.

To address this limitation, we need a framework for constructing resource-rational models in which both algorithmic and computational complexity can be systematically specified and manipulated. Meta-learning once again offers a principled framework to manipulate both algorithmic and computational complexity, and do so with relative ease. In Section 2.1 – Meta-learned models of cognition in Publications, I discuss how meta-learning makes it easy to manipulate the complexity – both computational and algorithmic – of the learning algorithm and explain previous work that has successfully utilized this capability. Additionally, in Section 2.3 – A resource-rational account of zero-shot compositional inference in a reinforcement learning setting in Publications, I demonstrate how a meta-learned model that accounts for limited algorithmic complexity best captures compositional reasoning in humans.

1.7 Meta-learning for ecologically-rational analysis

Another seminal framework, which, like rational analysis, emphasizes the role of environments in shaping behavior but eschews the optimality route it takes, is ecological rationality [11]. Ecological rationality proposes that humans use simple, resource-efficient heuristics tailored to environmental structure, enabling effective decision making in the real world even when parts of the information are unavailable and computational capacity is limited. It specifically "argues that environmentally successful algorithms can be developed without checking whether they approximate the optimal solution" [81], akin to rational analysis and resource-rational analysis. In doing so, it circumvents the computational intractability faced by the two previously discussed frameworks; see Section 1.5 and Section 1.6 in Background. Together, these factors make ecological rationality a flexible approach to explaining complex real-world behavior [8, 109].

The construction of models of ecological rationality begins with examining the structure of the decision environment, selecting heuristics from an adaptive toolbox (or hand designing one if none exists) and determining which are well-suited to the environment's properties. These heuristics are then evaluated based on how well they predict human behavior, with attention to trade-offs between accuracy, speed, and cognitive effort. The process is iterated by refining or introducing new heuristics until a satisfactory alignment with empirical data is achieved [8, 109]; see Box 11 for a detailed recipe used by ecological rationality.

11: The roots of ecological rationality lie in the notion of bounded rationality [108]. Simon argued that both the structure of the environment and the cognitive capacities of the agent are essential for explaining human rational behavior. Ecological rationality builds on this idea by proposing simple heuristics that not only conserve limited cognitive resources but also exploit specific features of the environment [8, 81, 109, 110].

[8]: Todd et al. (2012), *Ecological rationality: Intelligence in the world*
[109]: Gigerenzer et al. (2000), *Simple heuristics that make us smart*

Recipe prescribed for constructing models of rational analysis

1. **Characterize the environment**: Analyze the structure and properties of the decision environment
2. **Identify heuristics from the adaptive toolbox**: Select potential decision strategies from the collection of heuristics available
3. **Match heuristics to environments**: Determine which heuristics are well-suited to particular environments
4. **Test performance**: Evaluate how well the heuristics explain human behavior in different settings
5. **Study tradeoff**: Analyze the balance between accuracy, speed, and cognitive effort
6. **Reiterate**: If not satisfactory, introduce a new heuristic and reiterate.

The heuristics derived using the ecological rationality framework have successfully explained behaviors previously perceived as irrational, showing that in environments with skewed cue distributions, high uncertainty, or time pressure, people use simple strategies, such as take the best [111], recognition [112], or satisficing [97], which can rival or outperform more complex models [113, 114]. For instance, in tasks involving critical life decisions, such as choosing a partner, the importance of decision-relevant factors tends to decline rapidly [115]. That is, the most predictive cue is significantly more informative than the second, which in turn is more useful than the third, and so on. In such situations, people often use "fast and frugal" heuristics, particularly the *take-the-best* heuristic, which exploits this structure. This heuristic operates by ordering cues by their validity, searching them sequentially, and choosing the first cue that discriminates between alternatives, disregarding all remaining cues. Despite its simplicity, the performance of *take-the-best* rivals, and sometimes surpasses, that of more complex models such as multiple regression based on optimization techniques [113]. This example illustrates how simple heuristics, which leverage prior experience without aiming for formal optimality, can excel in ecological contexts [109, 113].

However, by taking this approach, ecological rationality sidesteps formal modeling in favor of hand-crafted heuristics that match specific environments. This is by design [16], as optimality-based approaches succeed in small world problems, where both the possible states the agent can take and their transitions are known, but fail to extend to large world problems, where states and transitions are unspecified among other unknown unknowns [116]. While this allows flexibility, it fails to provide a principled approach to derive heuristics. As a result, the current framework cannot explain how heuristics emerge, how they are chosen in a given context, how people switch between them when the properties of the environment change, or how prior experience shapes their selection and application [16, 87, 117]. This has resulted in the proliferation of task-specific heuristics, culminating in what is known as an adaptive toolbox [118, 119], without a unifying principle to structure them. Furthermore, traditional approaches to measuring real-world statistics in ecological rationality rely on data sources such as newspapers, magazines, online content, and curated surveys [8, 120]. These methods are often

time-consuming, expensive, and require extensive training and strict curation to mitigate sampling and demographic biases. As a result, they limit the scalability and broader applicability of ecological rationality as a general framework for explaining complex real-world behavior.

An alternative framework that offers a principled approach to derive heuristic-like strategies adapted to intricate structures present in real-world environments can help mitigate the shortcomings faced by ecological rationality. I will demonstrate that meta-learning can serve as this alternative framework. Specifically, I propose a novel methodology to derive models of ecological rationality using meta-learning, called ecological rational analysis. This method uses large-language models to extract tasks that capture rich statistics of the real world in multiple domains and meta-learning to derive a model optimally adapted to the extracted problems. In Section 2.4 – Human-like category learning by injecting ecological priors from large language models into neural networks in Publications, I show that LLMs can generate tasks, within the category learning domain, that capture real-world statistics and then show that meta-learning on those tasks results in an ecologically rational model which, in turn, explains human category learning both qualitatively and quantitatively. Later, in Section 2.5 – Meta-learning ecological priors from large language models captures human learning and decision making in Publications, I establish the domain-generality of the framework by showing that the models derived using ecological rational analysis capture human behavior in 15 experiments that span three different domains, namely, function learning, category learning, and decision making.

1.8 List of publications and contributions

In the following, I outline all the publications that I have contributed to over the course of my Ph.D. First, I list the subset of publications that are included in this thesis. They were chosen chosen to be in keeping with the focus of this thesis. After that, I list the other publications that I have contributed to, but which did not end up being part of this thesis.

1.8.1 Publications included in thesis

The list below contains the publications included in the thesis along with a detailed account of the contributions from various coauthors, including me. I describe contributions following the guidelines of the Contributor Roles Taxonomy (CRediT) [121], with minor modifications to ensure that the categories effectively capture the different components of a review article, an applied article, or a method article. Note that the symbol * denotes equal contribution.

- Binz, M., Dasgupta, I., **Jagadish, A. K.**, Botvinick, M., Wang, J. X., & Schulz, E. (2024). Meta-learned models of cognition. Behavioral and Brain Sciences, 47, e147.
  Initial idea: MB, AJ, ES
  Methodology and theoretical development: MB, ID, AJ, MBo, JW, and ES

Formal analysis: MB
Figures and visualization: MB
Writing (original draft): MB, ID, AJ, and ES
Writing (review and editing): MB, ID, AJ, MBo, JW, ES
Funding acquisition: ES

- Schubert, J. A., **Jagadish, A. K.**, Binz, M.*, & Schulz, E.* (2024). In-Context learning agents are asymmetric belief updaters. In Forty-first International Conference on Machine Learning.
Initial idea: AJ, MB
Methodology and theoretical development: JS, AJ, MB
Method validation and experiments: JS, AJ
Data preprocessing: JS
Figures and visualization: JS
Writing (original draft): JS, AJ, MB
Writing (review and editing): JS, AJ, MB, ES
Funding acquisition: ES

- **Jagadish, A. K.**, Binz, M., Saanum, T., Wang, J. X., Schulz, E. (2024). A resource-rational account of zero-shot compositional inference in a reinforcement learning setting.
Initial idea: AJ, ES
Methodology and theoretical development: AJ, MB, JW, ES
Method validation and experiments: AJ, MB, TS
Data collection and preprocessing: AJ
Figures and visualization: AJ
Writing (original draft): AJ, MB
Writing (review and editing): AJ, MB, TS, JW, ES
Funding acquisition: ES

- **Jagadish, A. K**., Coda-Forno, J., Thalmann, M., Schulz, E.*, & Binz, M.* (2024). Human-like category learning by injecting ecological priors from large language models into neural networks. In Forty-first International Conference on Machine Learning.
Initial idea: AJ, MB
Methodology and theoretical development: AJ, MB, MT, ES
Method validation and experiments: AJ, JCF, MB
Data collection and preprocessing: AJ
Figures and visualization: AJ
Writing (original draft): AJ, MB
Writing (review and editing): AJ, JCF, MT, ES, MB
Funding acquisition: ES

- **Jagadish, A. K**., Thalmann, M., Coda-Forno, J., Binz, M.*, & Schulz, E.* (2025). Meta-learning ecological priors from large language models captures human learning and decision making.
Initial idea: AJ, MB, ES
Methodology and theoretical development: AJ, MB, ES
Method validation and experiments: AJ, MB
Data collection and preprocessing: AJ
Figures and visualization: AJ
Writing (original draft): AJ
Writing (review and editing): AJ, JCF, MT, ES, MB

Funding acquisition: ES

1.8.2 Other publications

Below you find the list of publications that I contributed to during my Ph.D. that did not make the cut. For these publications, I only highlight my contributions.

- Rmus, M.*, **Jagadish, A. K.***, Mathony, M., Ludwig, T., & Schulz, E. (2025). Generating Computational Cognitive Models using Large Language Models. arXiv preprint arXiv:2502.00879.
  Contributions: AJ co-developed the methodology, conducted experiments to validate the methodology, preprocessed selected datasets, prepared selected figures, co-authored the original draft and edited the final manuscript.
- Cohen, Z.*, **Jagadish, A. K.***, Hosseini, E.*, & Eckstein, M. (2025). Reinforcement learning: Computational modeling of learning and decision-making. Transactions in Machine Learning Research (Under Review).
  Contributions: AJ contributed to to structuring the review, preparing selected figures, co-authoring the original draft and editing the final manuscript.
- Ben-Zion, Z., Witte, K.*, **Jagadish, A. K.***, Duek, O., Harpaz-Rotem, I., Khorsandian, M., ... & Spiller, T. R. (2025). 'Chat-GPT on the Couch': Assessing and Alleviating State Anxiety in Large Language Models. NPJ digital medicine.
  Contributions: AJ co-developed the method, conducted experiments to collect data, offered comments on figures and visualization, and edited the final manuscript.
- Binz, M., **Jagadish, A. K.**, Rmus, M., & Schulz, E. (2025). Automated scientific minimization of regret. arXiv:2505.17661.
  Contributions: AJ conducted experiments to provide a valid baseline for the proposed method, gave comments on the original draft and edited the final manuscript.
- Coda-Forno, J.*, Witte, K.*, **Jagadish, A. K.**, Binz, M., Akata, Z., & Schulz, E. (2024). Inducing anxiety in large language models increases exploration and bias. arXiv preprint arXiv:2304.11111.
  Contributions: AJ contributed to discussions about the methodology, figures and visualization, as well as writing (review and editing).
- Demircan, C.*, Saanum, T.*, **Jagadish, A. K.**, Binz, M., & Schulz, E. (2025). Sparse autoencoders reveal temporal difference learning in large language models. In Thirteenth International Conference on Learning Representations.
  Contributions: AJ contributed to discussions about the methodology, figures and visualization, as well as writing (review and editing).
- Binz, M., Akata, E., Bethge, M., Brändle, F., Callaway, F., Coda-Forno, J., ..., **Jagadish, A. K.**, ... & Schulz, E. (2024). Centaur: a foundation model of human cognition. Nature (arXiv preprint arXiv:2410.20268).

Contributions: AJ contributed to discussions about the methodology, implementation of baseline models, prompt design for selected experiments, figures and visualization, as well as writing (review and editing).

1.9 Organization of the remainder of the thesis

In the next part of the thesis, "How did it play out?", I present the publications that expand on how meta-learning can serve as a framework for building computational models of cognition and demonstrate how I have used it, specifically, to test theories of human learning. First, in Section 2.1 – Meta-learned models of cognition, I will illustrate how meta-learning enables the construction of Bayes-optimal learning algorithms, use this result to establish connection to rational analysis of cognition, and highlight its advantages over traditional modeling methods. Second, in Section 2.2 – In-context learning agents are asymmetric belief updaters, I will demonstrate how meta-learning can be used to perform rational analysis of optimistic belief updating displayed by both humans and large language models. Third, I will discuss how meta-learning offers the means to view zero-shot compositional inference in humans through the lens of resource-rationality; see Section 2.3 – A resource-rational account of zero-shot compositional inference in a reinforcement learning setting. Fourth, I will show that models of ecological rationality built using meta-learning can explain various aspects of human category learning; in Section 2.4 – Human-like category learning by injecting ecological priors from large language models into neural networks. Fifth, in Section 2.5 – Meta-learning ecological priors from large language models captures human learning and decision making, I will introduce the ecologically-rational analysis framework and show that models derived using this framework with meta-learning can explain human behavior in three cognitive domains, namely function learning, category learning, and decision making.

In the final part, "What did we learn?", I will contextualize these findings within the broader perspective outlined in the introduction, highlighting how the approach overcomes key challenges faced by traditional methods, and offers the means to unify rational analysis and its extensions. Broadly grouping them into four categories[12], I will detail the current limitations faced by meta-learned models of cognition, offer concrete proposals to address these limitations, and suggest directions for future research. Finally, I will discuss the potential implications of the framework for other fields and conclude with a concise summary of the main contributions of the thesis.

12: The four categories are: (1) Richer tasks and environments; (2) More plausible architectures, objectives and constraints; (3) Bridging the different levels of analysis; (4) Extending to other species.

How did it play out?

Publications 2

2.1 Meta-learned models of cognition

Binz, M., Dasgupta, I., **Jagadish, A. K.**, Botvinick, M., Wang, J. X., & Schulz, E. (2024). Meta-learned models of cognition. Behavioral and Brain Sciences, 47, e147. doi:10.1017/S0140525X23003266. arXiv:2304.06729.

Contributions in-context

The primary goal of this review article is to set up a research program around meta-learned models of cognition, which had been missing at the time of writing. To achieve this, we first contrasted how deriving computational models using meta-learning differs from traditional computational modeling approaches (see Section 1.3 in Background). After that, we illustrated how meta-learning can be used to construct Bayes-optimal learning algorithms by presenting a simplified version of the original formulation from Ortega et al., making their result more accessible to a broader audience. We then discussed a key implication of this result, namely, that it allows us to draw connections between meta-learning and the rational analysis of cognition (discussed in Section 1.5 in Background).

Later, we highlighted the key strengths of meta-learned models, explaining why it can serve as a replacement for Bayesian models at the center of rational analysis: First, meta-learning can produce approximately optimal learning algorithms when exact Bayesian inference is computationally intractable or when the inference problem cannot be phrased (see Section 2.2 in Publications for an empirical demonstration). Second, it allows straightforward manipulation of the computational and algorithmic complexity of the learning algorithms, enabling tests of resource-rational cognition (see also Section 1.6 in Background and Section 2.3 in Publications). Third, it facilitates the integration of insights from neuroscience into rational analysis through model architectures and objectives (see Section 3.2 in Outlook). In doing so, we demonstrated how the resulting models combine the strengths of two classic computational modeling frameworks, Bayesian [60] and connectionist models of cognition [64]; see Section 1.4 in Background.

Against this backdrop, we reviewed previous works, including some of my own, that have used meta-learning to study decision making heuristics, language understanding, acquisition of inductive biases, model-based reasoning, exploration and control. We concluded by addressing aspects of cognition that cannot be meta-learned, the role of neural networks in meta-learning, and a potential route towards building domain-general model of cognition.

Finally, in line with the standard protocol at *Behavioral and Brain Sciences*, we synthesized the commentaries written on the original piece and wrote a response article, organizing it into five sections: (1) data matters more

than we thought; (2) architecture matters, too; (3) transcending levels of analysis; (4) links to foundation models; (5) other points of contention (see Section 3.2 for an extended discussion).

Behavioral and Brain Sciences

cambridge.org/bbs

Meta-learned models of cognition

Marcel Binz[a,b], Ishita Dasgupta[c], Akshay K. Jagadish[a,b], Matthew Botvinick[c], Jane X. Wang[c] and Eric Schulz[a,b]

[a]Max Planck Institute for Biological Cybernetics, Tübingen, Germany; [b]Helmholtz Institute for Human-Centered AI, Munich, Germany and [c]Google DeepMind, London, UK
marcel.binz@helmholtz-munich.de
dasgupta.ishita@gmail.com
akshay.jagadish@tue.mpg.de
botvinick@google.com
wangjane@google.com
eric.schulz@tue.mpg.de

Target Article

Cite this article: Binz M, Dasgupta I, Jagadish AK, Botvinick M, Wang JX, Schulz E. (2024) Meta-learned models of cognition. *Behavioral and Brain Sciences* **47**, e147: 1–58. doi:10.1017/S0140525X23003266

Target Article Accepted: 19 November 2023
Target Article Manuscript Online: 23 November 2023
Commentaries Accepted: 13 February 2024

Keywords:
Bayesian inference; cognitive modeling; meta-learning; neural networks; rational analysis

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 20) and an Authors' Response (p. 53). See bbsonline.org for more information.

Corresponding author:
Marcel Binz;
Email: marcel.binz@helmholtz-munich.de

Abstract

Psychologists and neuroscientists extensively rely on computational models for studying and analyzing the human mind. Traditionally, such computational models have been hand-designed by expert researchers. Two prominent examples are cognitive architectures and Bayesian models of cognition. Although the former requires the specification of a fixed set of computational structures and a definition of how these structures interact with each other, the latter necessitates the commitment to a particular prior and a likelihood function that – in combination with Bayes' rule – determine the model's behavior. In recent years, a new framework has established itself as a promising tool for building models of human cognition: the framework of meta-learning. In contrast to the previously mentioned model classes, meta-learned models acquire their inductive biases from experience, that is, by repeatedly interacting with an environment. However, a coherent research program around meta-learned models of cognition is still missing to date. The purpose of this article is to synthesize previous work in this field and establish such a research program. We accomplish this by pointing out that meta-learning can be used to construct Bayes-optimal learning algorithms, allowing us to draw strong connections to the rational analysis of cognition. We then discuss several advantages of the meta-learning framework over traditional methods and reexamine prior work in the context of these new insights.

It is hard to imagine cognitive psychology and neuroscience without computational models – they are invaluable tools to study, analyze, and understand the human mind. Traditionally, such computational models have been hand-designed by expert researchers. In a cognitive architecture, for instance, researchers provide a fixed set of structures and a definition of how these structures interact with each other (Anderson, 2013b). In a Bayesian model of cognition, researchers instead specify a prior and a likelihood function that – in combination with Bayes' rule – fully determine the model's behavior (Griffiths, Kemp, & Tenenbaum, 2008). To provide one concrete example, consider the Bayesian model of function learning proposed by Lucas, Griffiths, Williams, and Kalish (2015). The goal of this model is to capture human learning in a setting that requires mapping input features to a numerical target value. When constructing their model, the authors had to hand-design a prior over functions that people expect to encounter. In this particular case, it was assumed that people prioritize linear functions over quadratic and other nonlinear functions.

The framework of meta-learning (Bengio, Bengio, & Cloutier, 1991; Schmidhuber, 1987; Thrun & Pratt, 1998) offers a radically different approach for constructing computational models by learning them through repeated interactions with an environment instead of requiring a priori specifications from a researcher. This process enables such models to acquire their inductive biases from experience, thereby departing from the traditional paradigm of hand-crafted models. For the function learning example mentioned above, this means that we do not need to specify which functions people expect to encounter in advance. Instead, during meta-learning a model would be exposed to many realistic function learning problems on which it then can figure out which functions are likely and which are not.

Recently, psychologists have started to apply meta-learning to the study of human learning (Griffiths et al., 2019). It has been shown that meta-learned models can capture a wide range of empirically observed phenomena that could not be explained otherwise. They, among others, reproduce human biases in probabilistic reasoning (Dasgupta, Schulz, Tenenbaum, & Gershman, 2020), discover heuristic decision-making strategies used by people (Binz, Gershman, Schulz, & Endres, 2022), and generalize compositionally on complex language tasks in a human-like manner (Lake & Baroni, 2023). The goal of the present article is to

develop a research program around meta-learned models of cognition and, in doing so, offer a synthesis of previous work and outline new research directions.

To establish such a research program, we will make use of a recent result from the machine learning community showing that meta-learning can be used to construct Bayes-optimal learning algorithms (Mikulik et al., 2020; Ortega et al., 2019; Rabinowitz, 2019). This correspondence is interesting from a psychological perspective because it allows us to connect meta-learning to another already well-established framework: the rational analysis of cognition (Anderson, 2013a; Chater & Oaksford, 1999). In a rational analysis, one first has to specify the goal of an agent along with a description of the environment the agent interacts with. The Bayes-optimal solution for the task at hand is then derived based on these assumptions and tested against empirical data. If needed, assumptions are modified and the whole process is repeated. This approach for constructing cognitive models has had a tremendous impact on psychology because it explains "why cognition works, by viewing it as an approximation to ideal statistical inference given the structure of natural tasks and environments" (Tenenbaum, 2021). The observation that meta-learned models can implement Bayesian inference implies that a meta-learned model can be used as a replacement for the corresponding Bayesian model in a rational analysis and thus suggests that any behavioral phenomenon that can be captured by a Bayesian model can also be captured by a meta-learned model.

We start our article by presenting a simplified version of an argument originally formulated by Ortega et al. (2019) and thereby make their result accessible to a broader audience. Having established that meta-learning produces models that can simulate Bayesian inference, we go on to discuss what additional explanatory power the meta-learning framework offers. After all, why should one not just stick to the tried-and-tested Bayesian approach? We answer this question by providing four original arguments in favor of the meta-learning framework (see Fig. 1 for a visual synopsis):

- Meta-learning can produce approximately optimal learning algorithms even if exact Bayesian inference is computationally intractable.
- Meta-learning can produce approximately optimal learning algorithms even if it is not possible to phrase the corresponding inference problem in the first place.
- Meta-learning makes it easy to manipulate a learning algorithm's complexity and can therefore be used to construct resource-rational models of learning.
- Meta-learning allows us to integrate neuroscientific insights into the rational analysis of cognition by incorporating these insights into model architectures.

The first two points highlight situations in which meta-learned models can be used for rational analysis but traditional Bayesian models cannot. The latter two points provide examples of how meta-learning enables us to make rational models of cognition more realistic, either by incorporating limited computational resources or neuroscientific insights. Taken together, these arguments showcase that meta-learning considerably extends the scope of rational analysis and thereby of cognitive theories more generally.

We will discuss each of these four points in detail and provide illustrations to highlight their relevance. We then reexamine prior studies from psychology and neuroscience that have applied meta-learning and put them into the context of our newly acquired insights. For each of the reviewed studies, we highlight how it relates to the four presented arguments, and discuss why its findings could not have been obtained using a classical Bayesian model. Following that, we describe under which conditions traditional models are preferable to those obtained by meta-learning. We finish our article by speculating what the future holds for meta-learning. Therein, we focus on how meta-learning could be the key to building a domain-general model of human cognition.

1. Meta-learned rationality

The prefix *meta-* is generally used in a self-referential sense: A meta-rule is a rule about rules, a meta-discussion is a discussion about discussions, and so forth. Meta-learning, consequently, refers to learning about learning. We, therefore, need to first establish a common definition of *learning* before covering meta-learning in more detail. For the present article, we adopt the following definition from Mitchell (1997):

Marcel Binz is a research scientist and deputy head of the Institute of Human-Centered AI at the Helmholtz Computational Health Center in Munich, and a guest scientist at the Max Planck Institute for Biological Cybernetics in Tübingen. He focuses on understanding human cognition from a computational perspective by utilizing tools from deep learning, reinforcement learning, Bayesian inference, and information theory. His work has been featured in top-tier journals such as *PNAS* and *Psychological Review*, as well as in leading machine learning venues such as NeurIPS and ICML.

Ishita Dasgupta is a senior research scientist at Google DeepMind. Her research is at the intersection of computational cognitive science and machine learning. She uses advances in machine learning to build new models of human reasoning, applies cognitive science approaches toward understanding black-box AI systems, and combines these insights to build more human-like artificial intelligence.

Akshay K. Jagadish is a PhD student jointly affiliated with the Max Planck Institute for Biological Cybernetics, Tübingen, and the Institute of Human-Centered AI at the Helmholtz Computational Health Center in Munich. His current research is dedicated to understanding the components essential for explaining human adaptive behavior across multiple task domains.

Matthew Botvinick, Senior Director of Research at Google DeepMind and Honorary Professor at Gatsby Unit for Computational Neuroscience, University College London, is the author of over 140 peer-reviewed publications in the areas of artificial intelligence, neuroscience, and cognitive psychology.

Jane X. Wang is a staff research scientist at Google DeepMind, where she researches meta-reinforcement learning, causal reasoning, and cognitively inspired AI. Her work has been published in various journals such as *Science*, *Neuron*, *Nature Neuroscience*, and top machine learning conferences such as NeurIPS, ICLR, and ICML.

Eric Schulz is the director of the Institute of Human-Centered AI at the Helmholtz Computational Health Center in Munich. He received his Ph.D. in Experimental Psychology from University College London in 2017. He is a recipient of the Robert J. Glushko Prize for Outstanding Doctoral Dissertation in Cognitive Science and a Jacobs Research Fellowship. His research focuses on understanding and improving human and machine learning.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

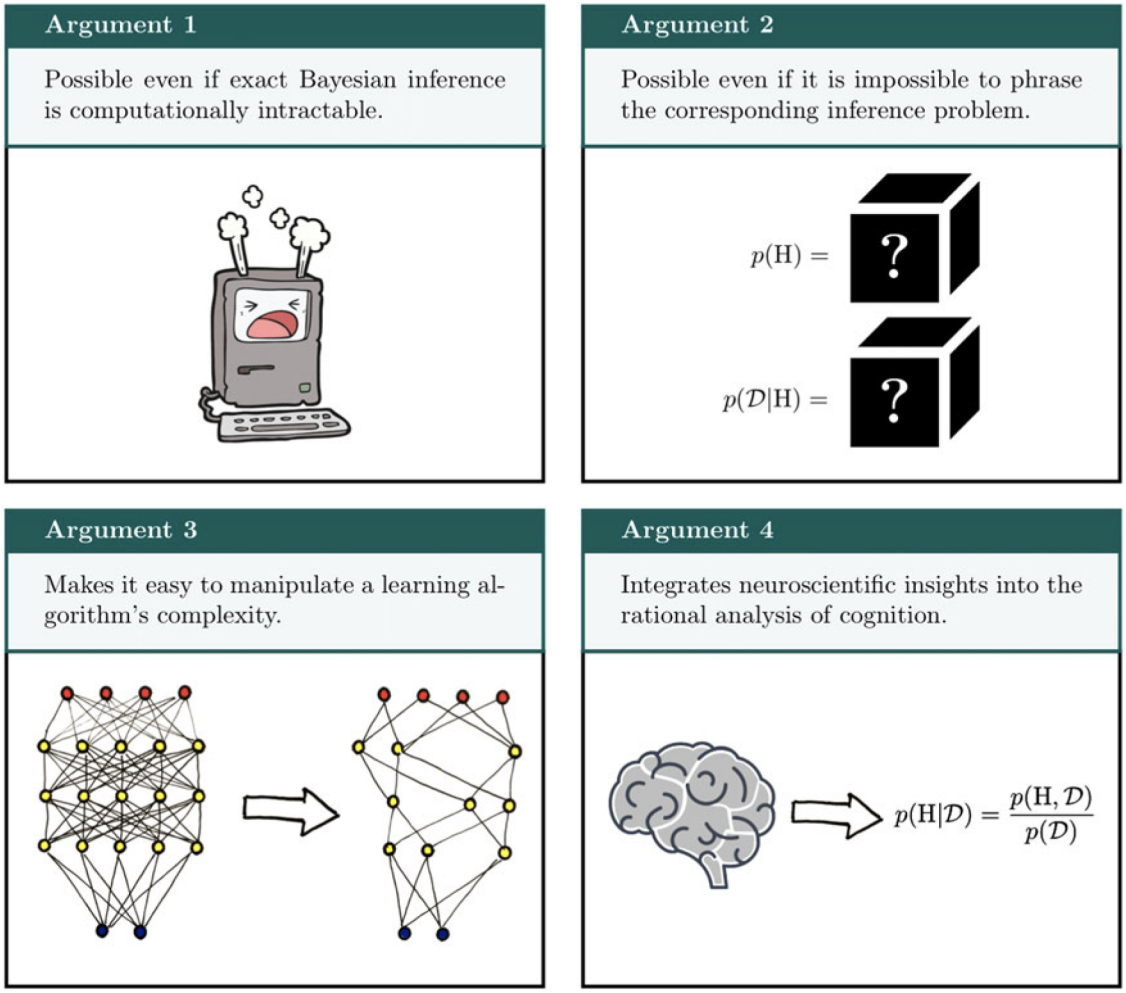

Figure 1. Visual synopsis of the four different arguments for meta-learning over Bayesian inference put forward in this article.

> **Definition: Learning**
> For a given task, training experience, and performance measure, an algorithm is said to learn if its performance at the task improves with experience.

To illustrate this definition, consider the following example which we will return to throughout the text: You are a biologist who has just discovered a new insect species and now set yourself the task of predicting how large members of this species are. You have already observed three exemplars in the wild with lengths of 16, 12, and 15 cm, respectively. These data amount to your training experience. Ideally, you can use this experience to make better predictions about the length of the next individual you encounter. You are said to have learned something if your performance is better after seeing the data than it was before. Typical performance measures for this example problem include the mean-squared error or the (negative) log-likelihood.

1.1 Bayesian inference for rational analyses

In a rational analysis of cognition, researchers are trying to compare human behavior to that of an optimal learning algorithm. However, it turns out that no learning algorithm is better than another when averaged over all possible problems (Wolpert, 1996; Wolpert & Macready, 1997), which means that we first have to make additional assumptions about the to-be-solved problem to obtain a well-defined notion of optimality. For our running example, one may make the following – somewhat unrealistic – assumptions:

(1) Each observed insect length x_k is sampled from a normal distribution with mean μ and standard deviation σ.
(2) An insect species' mean length μ cannot be observed directly, but the standard deviation σ is known to be 2 cm.
(3) Mean lengths across all insect species are distributed according to a normal distribution with a mean of 10 cm and a standard deviation of 3 cm.

An optimal way of making predictions about new observations under such assumptions is specified by Bayesian inference. Bayesian inference requires access to a prior distribution $p(\mu)$ that defines an agent's initial beliefs about possible parameter values before observing any data and a likelihood $p(x_{1:t}|\mu)$ that captures the agent's knowledge about how data are generated for a given set of parameters. In our running example, the prior and the likelihood can be identified as follows:

$$p(\mu) = N(\mu;\, 10,\, 3) \tag{1}$$

$$p(x_{1:t}|\mu) = \prod_{k=1}^{t} p\,(x_k|\mu) = \prod_{k=1}^{t} N(x_k;\, \mu,\, 2) \tag{2}$$

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

where $x_{1:t} = x_1, x_2, \ldots, x_t$ denote a sequence of observed insect lengths and the product in Equation (2) arises because of the additional assumption that observations are independent given the parameters.

The outcome of Bayesian inference is a posterior predictive distribution $p(x_{t+1}|x_{1:t})$, which the agent can use to make probabilistic predictions about a hypothetical future observation. To obtain this posterior predictive distribution, the agent first combines prior and likelihood into a posterior distribution over parameters by applying Bayes' theorem:

$$p(\mu|x_{1:t}) = \frac{p(x_{1:t}|\mu)p(\mu)}{\int p(x_{1:t}|\mu)p(\mu)d\mu} \tag{3}$$

In a subsequent step, the agent then averages over all possible parameter values weighted by their posterior probability to get the posterior predictive distribution:

$$p(x_{t+1}|x_{1:t}) = \int p(x_{t+1}|\mu)p(\mu|x_{1:t})d\mu \tag{4}$$

Multiple arguments justify Bayesian inference as a normative procedure, and thereby its use for rational analyses (Corner & Hahn, 2013). This includes Dutch book arguments (Lewis, 1999; Rescorla, 2020), free-energy minimization (Friston, 2010; Hinton & Van Camp, 1993), and performance-based justifications (Aitchison, 1975; Rosenkrantz, 1992). For this article, we are mainly interested in the latter class of performance-based justifications because these can be used – as we will demonstrate later on – to derive meta-learning algorithms that learn approximations to Bayesian inference.

Performance-based justifications are based on the notion of frequentist statistics. They assert that no learning algorithm can be better than Bayesian inference on a certain performance measure. Particularly relevant for this article is a theorem first proved by Aitchison (1975). It states that the posterior predictive distribution is the distribution from the set of all possible distributions Q that maximizes the log-likelihood of hypothetical future observations when averaged over the data-generating distribution $p(\mu, x_{1:t+1}) = p(\mu)p(x_{1:t+1}|\mu)$:

$$p(x_{t+1}|x_{1:t}) = \text{argmax}_{q \in Q} E_{p(\mu, x_{1:t+1})}[\log q(x_{t+1}|x_{1:t})] \tag{5}$$

Equation (5) implies that if an agent wants to make a prediction about the length of a still unobserved exemplar from a particular insect species and measures its performance using the log-likelihood, then – averaged across all possible species that can be encountered – there is no better way of doing it than using the posterior predictive distribution. We decided to include a short proof of this theorem within Box 1 for the curious reader as it does not appear in popular textbooks on probabilistic machine learning (Bishop, 2006; Murphy, 2012) nor in survey articles on Bayesian models of cognition. Note that, although the theorem itself is central to our later argument, working through its proof is not required to follow the remainder of this article.

Box 1 Proof: meta-learning maximizes log-likelihoods of future observations

We proof that the posterior predictive distribution $p(x_{t+1}|x_{1:t})$ maximizes the log-likelihood of future observations averaged over the data-generating distribution:

$$p(x_{t+1}|x_{1:t}) = \text{argmax}_q E_{p(\mu, x_{1:t+1})}[\log q(x_{t+1}|x_{1:t})] \tag{8}$$

The essence of this proof is to show that the posterior predictive distribution is superior to any other reference distribution $r(x_{t+1}|x_{1:t})$ in terms of log-likelihood:

$$E_{p(\mu, x_{1:t})}[\log p(x_{t+1}|x_{1:t})] \geq E_{p(\mu, x_{1:t})}[\log r(x_{t+1}|x_{1:t})]$$

or equivalently that:

$$E_{p(\mu, x_{1:t})}\left[\log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})}\right] \geq 0$$

Proofing this conjecture is straight-forward (Aitchison, 1975):

$$\begin{aligned}
&E_{p(\mu, x_{1:t})}\left[\log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})}\right] \\
&= \sum_{\mu}\sum_{x_{1:t}}\sum_{x_{t+1}} \log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})} p(x_{t+1}|\mu)p(x_{1:t}|\mu)p(\mu) \\
&= \sum_{x_{1:t}}\sum_{\mu}\sum_{x_{t+1}} \log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})} p(x_{t+1}|\mu)p(x_{1:t}|\mu)p(\mu) \\
&= \sum_{x_{1:t}}\sum_{\mu}\sum_{x_{t+1}} \log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})} p(x_{t+1}|\mu)p(\mu|x_{1:t})p(x_{1:t}) \\
&= \sum_{x_{1:t}}\left[\sum_{\mu}\sum_{x_{t+1}} \log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})} p(x_{t+1}|\mu)p(\mu|x_{1:t})\right]p(x_{1:t}) \\
&= \sum_{x_{1:t}}\left[\sum_{x_{t+1}}\sum_{\mu} \log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})} p(x_{t+1}|\mu)p(\mu|x_{1:t})\right]p(x_{1:t}) \\
&= \sum_{x_{1:t}}\left[\sum_{x_{t+1}} \log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})}\left[\sum p(x_{t+1}|\mu)p(\mu|x_{1:t})\right]\right]p(x_{1:t}) \\
&= \sum_{x_{1:t}}\left[\sum_{x_{t+1}} \log \frac{p(x_{t+1}|x_{1:t})}{r(x_{t+1}|x_{1:t})} p(x_{t+1}|x_{1:t})\right]p(x_{1:t}) \\
&= \sum_{x_{1:t}} KL\left[p(x_{t+1}|x_{1:t})|r(x_{t+1}|x_{1:t})\right]p(x_{1:t}) \\
&\geq 0
\end{aligned}$$

Note that although we used sums in our proof, thereby assuming that relevant quantities take discrete values, the same ideas can be readily applied to continuous-valued quantities by replacing sums with integrals.

1.2 Meta-learning

Having summarized the general concepts behind Bayes-optimal learning, we can now start to describe meta-learning in more detail. Formally speaking, a meta-learning algorithm is defined as any algorithm that "uses its experience to change certain aspects of a learning algorithm, or the learning method itself, such that the modified learner is better than the original learner at learning from additional experience" (Schaul & Schmidhuber, 2010).

To accomplish this, one first decides on an inner-loop (or base) learning algorithm and determines which of its aspects can be modified. We also refer to these modifiable aspects as meta-parameters (i.e., meta-parameters are simply parameters of a system that are adapted during meta-learning). In an outer-loop (or meta-learning) process, the system is then trained on a series of learning problems such that the inner-loop learning algorithm gets better at solving the problems that it encounters. We provide a high-level overview of this framework in Figure 2.

The previous definition is quite broad and includes a variety of methods. It is, for example, possible to meta-learn:

- Hyperparameters for a base learning algorithm, such as learning rates, batch sizes, or the number of training epochs

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

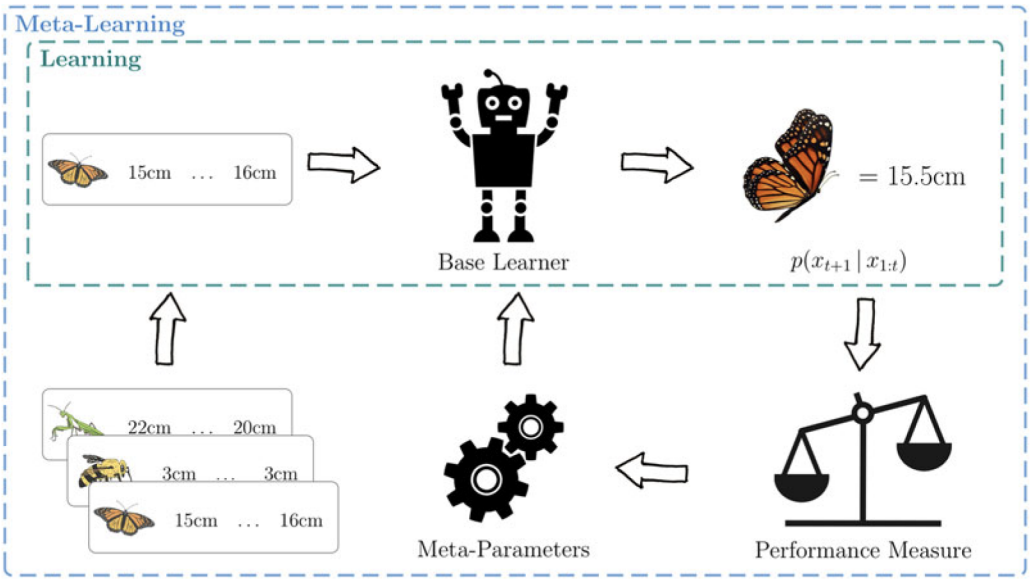

Figure 2. High-level overview of the meta-learning process. A base learner (green rectangle) receives data and performs some internal computations that improve its predictions on future data-points. A meta-learner (blue rectangle) encompasses a set of meta-parameters that can be adapted to create an improved learner. This is accomplished by training the learner on a distribution of related learning problems.

(Doya, 2002; Feurer & Hutter, 2019; Li, Zhou, Chen, & Li, 2017).

- Initial parameters of a neural network that is trained via stochastic gradient descent (Finn, Abbeel, & Levine, 2017; Nichol, Achiam, & Schulman, 2018).
- Prior distributions in a probabilistic graphical model (Baxter, 1998; Grant, Finn, Levine, Darrell, & Griffiths, 2018).
- Entire learning algorithms (Hochreiter, Younger, & Conwell, 2001; Santoro, Bartunov, Botvinick, Wierstra, & Lillicrap, 2016).

Although all these methods have their own merits, we will be primarily concerned with the latter approach. Learning entire learning algorithms from scratch is arguably the most general and ambitious type of meta-learning, and it is the focus of this article because it is the only one among the aforementioned approaches leading to Bayes-optimal learning algorithms that can be used for rational analyses.

1.3 Meta-learned inference

It may seem like a daunting goal to learn an entire learning algorithm from scratch, but the core idea behind the approach we discuss in the following is surprisingly simple: Instead of using Bayesian inference to obtain the posterior predictive distribution, we teach a general-purpose function approximator to do this inference. Previous work has mostly focused on using recurrent neural networks as function approximators in this setting and thus we will – without loss of generality – focus our upcoming exposition on this class of models.

Like the posterior predictive distribution, the recurrent neural network processes a sequence of observed length from a particular insect species and produces a predictive distribution over the lengths of potential future observations from the same species. More concretely, the meta-learned predictive distribution takes a predetermined functional form whose parameters are given by the network outputs. If we had, for example, decided to use a normal distribution as the functional form of the meta-learned predictive distribution, outputs of the network would correspond to an expected length m_{t+1} and its standard deviation s_{t+1}. Figure 3a illustrates this setup graphically.

Initially, the recurrent neural network implements a randomly initialized learning algorithm.[1] The goal of the meta-learning process is then to turn this system into an improved learning algorithm. The final result is a learning algorithm that is *learned* or trained rather than specified by a practitioner. To create a learning signal to do this training, we need a performance measure that can be used to optimize the network. Equation (5) suggests a straightforward strategy for designing such a measure by replacing the maximization over all possible distributions with a maximization over meta-parameters Θ (in our case, the weights of the recurrent neural network):

$$\begin{aligned} &\operatorname{argmax}_{q \in Q} E_{p(\mu, x_{1:t+1})}[\log q\,(x_{t+1} | x_{1:t})] \\ &\quad \approx \operatorname{argmax}_{\Theta} E_{p(\mu, x_{1:t+1})}[\log q\,(x_{t+1} | x_{1:t}, \Theta)] \end{aligned} \tag{6}$$

To turn this expression into a practical meta-learning algorithm, we will – as common practice when training deep neural networks – maximize a sample-based version using stochastic gradient ascent:

$$\begin{aligned} &\operatorname{argmax}_{\Theta} E_{p(\mu, x_{1:t+1})}[\log q\,(x_{t+1} | x_{1:t}, \Theta)] \\ &\quad \approx \operatorname{argmax}_{\Theta} \frac{1}{N} \sum_{n=1}^{N} \log q\,(x_{t+1}^{(n)} | x_{1:t}^{(n)}, \Theta) \end{aligned} \tag{7}$$

Figure 3b presents pseudocode for a simple gradient-based procedure that maximizes Equation (7). The entire meta-learning algorithm can be implemented in just around 30 lines of self-contained *PyTorch* code (Paszke et al., 2019). We provide an annotated reference implementation on this article's accompanying Github repository.[2]

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

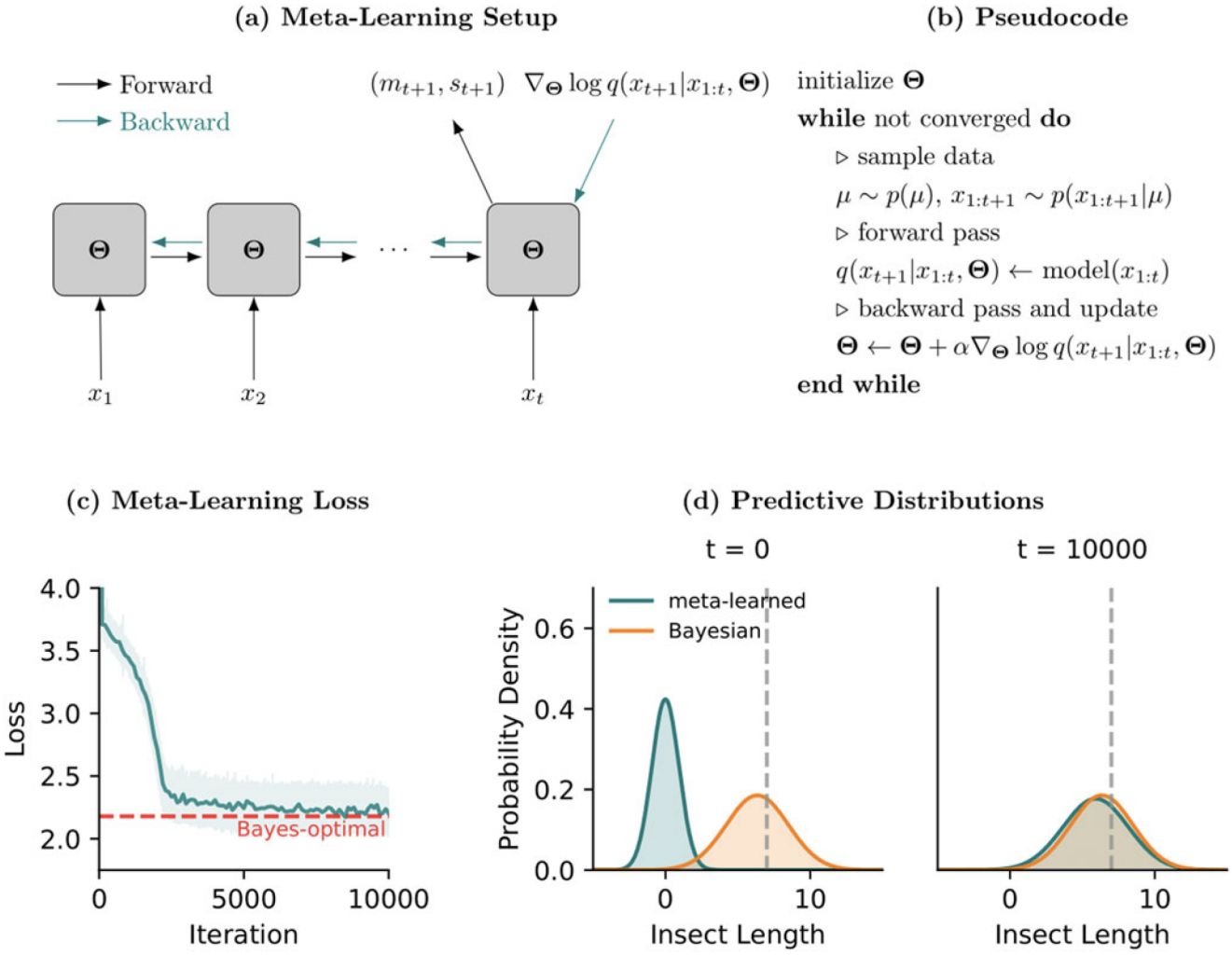

Figure 3. Meta-learning illustration. (a) A recurrent neural network processes a sequence of observations and produces a predictive distribution at the final time-step. (b) Pseudocode for a simple meta-learning algorithm. (c) Loss during meta-learning with shaded contours corresponding to the standard deviation across 30 runs. (d) Posterior and meta-learned predictive distributions for an example sequence at beginning and end of meta-learning. The dotted gray line denotes the (unobserved) mean length.

1.4 How good is a meta-learned algorithm?

We have previously shown that the global optimum of Equation (7) is achieved by the posterior predictive distribution. Thus, by maximizing this performance measure, the network is actively encouraged to implement an approximation to exact Bayesian inference. Importantly, after meta-learning is completed, producing an approximation to the posterior predictive distribution does not require any further updates to the network weights. To perform an inference (i.e., the learn), we simply have to query the network's outputs after providing it with a particular sequence of observations. Learning at this stage is then realized by updating the hidden activations of the recurrent neural network as opposed to its weights. The characteristics of this new activation-based learning algorithm can be potentially vastly different from the weight-based learning algorithm used for meta-learning.

If we want to use the fully optimized network for rational analyses, we have to ask ourselves: How well does the resulting model approximate Bayesian inference? Two aspects have to be considered when answering this question. First, the network has to be sufficiently expressive to produce the exact posterior predictive distribution for all input sequences. Neural networks of sufficient width are universal function approximators (Hornik, Stinchcombe, & White, 1989), meaning that they can approximate any continuous function to arbitrary precision. Therefore, this aspect is not too problematic for the optimality argument. The second aspect is a bit more intricate: Assuming that the network is powerful enough to represent the global optimum of Equation (7), the employed optimization procedure also has to find it. Although we are not aware of any theorem that could provide such a guarantee, in practice, it has been observed that meta-learning procedures similar to the one discussed here often lead to networks that closely approximate Bayesian inference (Mikulik et al., 2020; Rabinowitz, 2019). We provide a visualization demonstrating that the predictions of a meta-learned model closely resemble those of exact Bayesian inference for our insect length example in Figures 3c and 3d.

Although our exposition in this section focused on the supervised learning case, the same ideas can also be readily extended to the reinforcement learning setting (Duan et al., 2016; Wang et al., 2016). Box 2 outlines the general ideas behind the meta-reinforcement learning framework.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Box 2 Meta-reinforcement learning

The main text has focused on tasks in which an agent receives direct feedback about which response would have been correct. In the real world, however, people do not always receive such explicit feedback. They, instead, often have to deal with partial information – taking the form of rewards, utilities, or costs – that merely informs them about the quality of their response.

Problems that fall into this category are often modeled as Markov decision processes (MDPs). In an MDP, an agent repeatedly interacts with an environment. In each time-step, it observes the state of the environment s_t and can take an action a_t that leads to a reward signal r_t sampled from a reward distribution $p(r_t|s_t, a_t, \mu_r)$. Executing an action furthermore influences the environment state at the next time-step according to a transition distribution $p(s_{t+1}|s_t, a_t, \mu_s)$.

The goal of a Bayes-optimal reinforcement learning agent is to find a policy, which is a mapping from a history of observations $h_t = s_1, a_1, r_1, \ldots, s_{t-1}, a_{t-1}, r_{t-1}, s_t$ to a probability distribution over actions $\pi^*(a_t|h_t)$, that maximizes the total amount of obtained rewards across a finite horizon H averaged over all problems that may be encountered:

$$\pi^*(a_t|h_t) = \text{argmax}_\pi E_{p(\mu_r,\mu_s)} \prod p(r_t|s_t,a_t,\mu_r) p(s_{t+1}|s_t,a_t,\mu_s) \pi(a_t|h_t) \left[\sum_{t=1}^{H} r_t \right] \quad (9)$$

MDPs with unknown reward and transition distributions are substantially more challenging to solve optimally compared to supervised problems as there is no teacher informing the agent about which actions are right or wrong. Instead, the agent has to figure out the most rewarding course of action solely through trial and error. Finding an analytical solution to Equation (9) is extremely challenging and indeed only possible for a few special cases (Duff, 2003; Gittins, 1979), which made it historically near impossible to investigate such problems within the framework of rational analysis.

Even though finding an analytical expression of the Bayes-optimal policy is often impossible, it is straightforward to meta-learn an approximation to it (Duan et al., 2016; Wang et al., 2016). The general concept is almost identical to the supervised learning case: Parameterize the to-be-learned policy with a recurrent neural network and replace the maximization over the set of all possible policies from Equation (9) with a sample-based approximation that maximizes over neural network parameters. The resulting problem can then be solved using any standard deep reinforcement learning algorithm.

Like in the supervised learning case, the resulting recurrent neural network implements a free-standing reinforcement learning algorithm after meta-learning is completed. Learning is once again implemented via a simple forward pass through the network, i.e., by conditioning the model on an additional data-point. The meta-learned reinforcement learning algorithm approximates the Bayes-optimal policy under the same conditions as in the supervised learning case: A sufficiently expressive model and an optimization procedure that is able to find the global optimum.

1.5 Tool or theory?

It is often not so trivial to separate meta-learning from normal learning. We believe that part of this confusion arises from an underspecification regarding what is being studied. In particular, the meta-learning framework provides opportunities to address two distinct research questions:

(1) It can be used to study how people improve their learning abilities over time.
(2) It can be used as a methodological tool to construct learning algorithms with the properties of interest (and thereafter compare the emerging learning algorithms to human behavior).

Historically, behavioral psychologists have been mainly interested in the former aspect (Doya, 2002; Harlow, 1949). In the 1940s, for example, Harlow (1949) already studied how learning in monkeys improves over time. He found that they adapted their learning strategies after sufficiently many interactions with tasks that shared a common structure, thereby showing a learning-to-learn effect. By now, examples of this phenomenon have been found in many different species – including humans – across nature (Wang, 2021).

More recently, psychologists have started to view meta-learning as a methodological tool to construct approximations to Bayes-optimal learning algorithms (Binz et al., 2022; Kumar, Dasgupta, Cohen, Daw, & Griffiths, 2020a), and subsequently use the resulting algorithms to study human cognition. The key difference from the former approach is that, in this setting, one abstracts away from the process of meta-learning and instead focuses on its outcome. From this perspective, only the fully converged model is of interest. Importantly, this approach allows us to investigate human learning from a rational perspective because we have demonstrated that meta-learning can be used to construct approximations to Bayes-optimal learning.

We place an emphasis on the second aspect in the present article and advocate for using fully converged meta-learned algorithms – as replacements for the corresponding Bayesian models – for rational analyses of cognition.[3] In the next section, we will outline several arguments that support this approach. However, it is important to mention that we believe that meta-learning can also be a valuable tool to understand the process of learning-to-learn itself. In this context, several intriguing questions arise: At what timescale does meta-learning take place in humans? How much of it is because of task-specific adaptations? How much of it is based on evolutionary or developmental processes? Although we agree that these are important questions, they are not the focus of this article.

2. Why not Bayesian inference?

We have just argued that it is possible to meta-learn Bayes-optimal learning algorithms. What are the implications of this result? If one has access to two different theories that make identical predictions, which of them should be preferred? Bayesian inference has already established itself as a valuable tool for building cognitive models in the recent decades. Thus, the burden of proof is arguably on the meta-learning framework. In this section, we provide four different arguments that highlight the advantages of meta-learning for building models of cognition. Many of these arguments are novel and have not been put forward explicitly in previous literature. The first two arguments highlight situations in which meta-learned models can be used for rational analysis but traditional Bayesian models cannot. The latter two provide examples of how meta-learning enables us to make rational models of cognition more realistic, either by incorporating limited computational resources or neuroscientific insights.

2.1 Intractable inference

Argument 1
Meta-learning can produce approximately optimal learning algorithms even if exact Bayesian inference is computationally intractable.

Bayesian inference becomes intractable very quickly because the complexity of computing the normalization constant that appears in the denominator grows exponentially with the number of unobserved parameters. In addition, it is only possible to find a closed-form expression of the posterior distribution for certain combinations of prior and likelihood. In our running example, we assumed that both prior and likelihood follow a normal distribution, which, in turn, leads to a normally distributed posterior. However, if one would instead assume that the prior over mean

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

length follows an exponential distribution – which is arguably a more sensible assumption as it enforces lengths to be positive – it becomes already impossible to find an analytical expression for the posterior distribution.

Researchers across disciplines have recognized these challenges and have, in turn, developed approaches that can approximate Bayesian inference without running into computational difficulties. Prime examples of this are variational inference (Jordan, Ghahramani, Jaakkola, & Saul, 1999) and Markov chain Monte-Carlo (MCMC) methods (Geman & Geman, 1984). In variational inference, one phrases inference as an optimization problem by positing a variational approximation whose parameters are fitted to minimize a divergence measure to the true posterior distribution. MCMC methods, on the other hand, draw samples from a Markov chain that has the posterior distribution as its equilibrium distribution. Previous research in cognitive science indicates that human learning shows characteristics of such approximations (Courville & Daw, 2008; Dasgupta, Schulz, & Gershman, 2017; Daw, Courville, & Dayan, 2008; Sanborn, Griffiths, & Navarro, 2010; Sanborn & Silva, 2013).

Meta-learned inference also never requires an explicit calculation of the exact posterior or posterior predictive distribution. Instead, it performs approximately optimal inference via a single forward pass through the network. Inference, in this case, is approximate because we had to determine a functional form for the predictive distribution. The chosen form may deviate from the true form of the posterior predictive distribution, which, in turn, leads to approximation errors.[4] In some sense, this type of approximation is similar to variational inference: Both approaches involve optimization and require one to define a functional form of the respective distribution. However, the optimization process in both approaches uses a different loss function and happens at different timescales. Although variational inference employs the negative evidence lower bound as its loss function, meta-learning directly maximizes for models that can be expected to generalize well to unseen observations (using the performance-based measure from Equation (5)). Furthermore, meta-learned inference only involves optimization during the outer-loop meta-learning process but not during the actual learning itself. To update how a meta-learned model makes predictions in the light of new data, we only have to perform a simple forward pass through the network. In contrast to this, standard variational inference requires us to rerun the whole optimization process from scratch every time a new data-point is observed.[5]

In summary, it is possible to meta-learn an approximately Bayes-optimal learning algorithm. If exact Bayesian inference is not tractable, such models are our best option for performing rational analyses. Yet, many other methods for approximate inference, such as variational inference and MCMC methods, also share this feature, and it will thus ultimately be an empirical question which of these approximations provides a better description of human learning.

2.2 Unspecified problems

Argument 2
Meta-learning can produce optimal learning algorithms even if it is not possible to phrase the corresponding inference problem in the first place.

Bayesian inference is hard, but posing the correct inference problem can be even harder. What exactly do we mean by that? To perform Bayesian inference, we need to specify a prior and a likelihood. Together, these two objects fully specify the assumed data-generating distribution, and thus the inference problem. Ideally, the specified data-generating distribution should match how the environment actually generates its data. It is fairly straightforward to fulfill this requirement in artificial scenarios, but for many real-world problems, it is not. Take for instance our running example: Does the prior over mean length really follow a normal distribution? If yes, what are the mean and variance of this distribution? Are the underlying parameters actually time-invariant or do they, for example, change based on seasons? None of these questions can be answered with certainty.

In his seminal work on Bayesian decision theory, Savage (1972) made the distinction between small- and large-world problems. A small-world problem is one "in which all relevant alternatives, their consequences, and probabilities are known" (Gigerenzer & Gaissmaier, 2011). A large-world problem, on the other hand, is one in which the prior, the likelihood, or both cannot be identified. Savage's distinction between small and large worlds is relevant for the rational analysis of human cognition as its critics have pointed out that Bayesian inference only provides a justification for optimal reasoning in small-world problems (Binmore, 2007) and that "very few problems of interest to the cognitive, behavioral, and social sciences can be said to satisfy [this] condition" (Brighton & Gigerenzer, 2012).

Identifying the correct set of assumptions becomes especially challenging once we deal with more complex problems. To illustrate this, consider a study conducted by Lucas et al. (2015) who attempted to construct a Bayesian model of human function learning. Doing so required them to specify a prior over functions that people expect to encounter. Without direct access to such a distribution, they instead opted for a heuristic solution: 98.8% of functions are expected to be linear, 1.1% are expected to be quadratic, and 0.1% are expected to be nonlinear. Empirically, this choice led to good results, but it is hard to justify from a rational perspective. We simply do not know the frequency with which these functions appear in the real world, nor whether the given selection fully covers the set of functions expected by participants.

There are also inference problems in which it is not possible to specify or compute the likelihood function. These problems have been studied extensively in the machine learning community under the names of simulation-based or likelihood-free inference (Cranmer, Brehmer, & Louppe, 2020; Lueckmann, Boelts, Greenberg, Goncalves, & Macke, 2021). In this setting, it is typically assumed that we can sample data from the likelihood for a given parameter setting but that computing the corresponding likelihood is impossible. Take, for instance, our insect length example. It should be clear that an insect's length does not only depend on its species' mean but also on many other factors such as climate, genetics, and the individual's age. Even if all these factors were known, mapping them to a likelihood function does seem close to impossible.[6] But, we can generate samples easily by observing insects in the wild. If we had access to large database of insect length measurements for different species, this could be directly used to meta-learn an approximately Bayes-optimal learning algorithm for predicting their length, while circumventing an explicit definition of a likelihood function.

In cases where we do not have access to a prior or a likelihood, we can neither apply exact Bayesian inference nor approximate inference schemes such as variational inference or MCMC methods. In contrast to this, meta-learned inference does not require us to define the prior or the likelihood explicitly. It only demands samples from the data-generating distribution to meta-learn an

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

approximately Bayes-optimal learning algorithm – a much weaker requirement (Müller, Hollmann, Arango, Grabocka, & Hutter, 2021). The ability to construct Bayes-optimal learning algorithms for large-world problems is a unique feature of the meta-learning framework, and we believe that it could open up totally new avenues for constructing rational models of human cognition. To highlight one concrete example, it would be possible to take a collection of real-world decision-making tasks – such as the ones presented by Czerlinski et al. (1999) – and use them to obtain a meta-learned agent that is adapted to the decision-making problems that people actually encounter in their everyday lives. This algorithm could then serve as a normative standard against which we can compare human decision making.

2.3 Resource rationality

Argument 3
Meta-learning makes it easy to manipulate a learning algorithm's complexity and can therefore be used to construct resource-rational models of learning.

Bayesian inference has been successfully applied to model human behavior across a number of domains, including perception (Knill & Richards, 1996), motor control (Körding & Wolpert, 2004), everyday judgments (Griffiths & Tenenbaum, 2006), and logical reasoning (Oaksford et al., 2007). Notwithstanding these success stories, there are also well-documented deviations from the notion of optimality prescribed by Bayesian inference. People, for example, underreact to prior information (Kahneman & Tversky, 1973), ignore evidence (Benjamin, 2019), and rely on heuristic decision-making strategies (Gigerenzer & Gaissmaier, 2011).

The intractability of Bayesian inference – together with empirically observed deviations from it – has led researchers to conjecture that people only attempt to approximate Bayesian inference. Many different notions of what constitutes a computational resource have been suggested, such as memory (Dasgupta & Gershman, 2021), thinking time (Ratcliff & McKoon, 2008), or physical effort (Hoppe & Rothkopf, 2016).

Cover (1999) relies on a dichotomy that will be useful for our following discussion. He refers to the algorithmic complexity of an algorithm as the number of bits needed to *implement* it. In contrast, he refers to the computational complexity of an algorithm as the space, time, or effort required to *execute* it. It is possible to cast many approximate inference schemes as resource-rational algorithms (Sanborn, 2017). The resulting models typically consider some form of computational complexity. In MCMC methods, computational complexity can be measured in terms of the number of drawn samples: Drawing fewer samples leads to faster inference at the cost of introducing a bias (Courville & Daw, 2008; Sanborn et al., 2010). In variational inference, on the other hand, it is possible to introduce an additional parameter that allows to trade-off performance against the computational complexity of transforming the prior into the posterior distribution (Binz & Schulz, 2022b; Ortega, Braun, Dyer, Kim, & Tishby, 2015). Likewise, other frameworks for building resource-rational models, such as rational metareasoning (Lieder & Griffiths, 2017), also only target computational complexity.

The prevalence of resource-rational models based on computational complexity is likely because of the fact that building similar models based on algorithmic complexity is much harder. Measuring algorithmic complexity historically relies on the notion of Kolmogorov complexity, which is the size of the shortest computer program that produces a particular data sequence (Chaitin, 1969; Kolmogorov, 1965; Solomonoff, 1964). Kolmogorov complexity is in general noncomputable, and, therefore, of limited practical interest.[7]

Meta-learning provides us with a straightforward way to manipulate both algorithmic and computational complexity in a common framework by adapting the size of the underlying neural network model. Limiting the complexity of network weights places a constraint on algorithmic complexity (as reducing the number of weights decreases the number of bits needed to store them, and hence also the number of bits needed to store the learning algorithm). Limiting the complexity of activations, on the other hand, places a constraint on computational complexity (reducing the number of hidden units, e.g., decreases the memory needed for executing the meta-learned model). This connection can be made more formal in an information-theoretic framework (Hinton & Van Camp, 1993; Hinton & Zemel, 1993). For applications of this idea in the context of human cognition, see, for instance, Binz et al. (2022) or Bates and Jacobs (2020).

Previously, both forms of complexity constraints have been realized in meta-learned models. Dasgupta et al. (2020) decreased the number of hidden units of a meta-learned inference algorithm, effectively reducing its computational complexity. In contrast, Binz et al. (2022) placed a constraint on the description length of neural network weights (i.e., the number of bits required to store them), which implements a form of algorithmic complexity. To the best of our knowledge, no other class of resource-rational models exists that allows us to take both algorithmic and computational complexity into account, making this ability a unique feature of the meta-learning framework.

2.4 Neuroscience

Argument 4
Meta-learning allows us to integrate neuroscientific insights into the rational analysis of cognition by incorporating these insights into model architectures.

In addition to providing a framework for understanding many aspects of behavior, meta-learning offers a powerful lens through which to view brain structure and function. For instance, Wang et al. (2018) presented observations supporting the hypothesis that prefrontal circuits may constitute a meta-reinforcement learning system. From a computational perspective, meta-learning strives to learn a faster inner-loop learning algorithm via an adjustment of neural network weights in a slower outer-loop learning process. Within the brain, an analogous process plausibly occurs when slow, dopamine-driven synaptic change gives rise to reinforcement learning processes that occur within the activity dynamics of the prefrontal network, allowing for adaptation on much faster timescales. This perspective recontextualized the role of dopamine function in reward-based learning and was able to account for a range of previously puzzling neuroscientific findings. To highlight one example, Bromberg-Martin, Matsumoto, Hong, and Hikosaka (2010) found that dopamine signaling reflected updates in not only *experienced* but also *inferred* values of targets. Notably, a meta-reinforcement learning agent trained on the same task also recovered this pattern. Having a mapping of meta-reinforcement learning components onto existing brain regions furthermore allows us to apply experimental manipulations that directly perturb neural activity, for example by using optogenetic techniques. Wang et al. (2018) used this idea to

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

modify their original meta-reinforcement learning architecture to mimic the blocking or enhancement of dopaminergic reward prediction error signals, in direct analogy with optogenetic stimulation delivered to rats performing a two-armed bandit task (Stopper, Maric, Montes, Wiedman, & Floresco, 2014).

Importantly, the direction of exchange can also work in the other direction, with neuroscientific findings constraining and inspiring new forms of meta-learning architectures. Bellec, Salaj, Subramoney, Legenstein, and Maass (2018), for example, showed that recurrent networks of spiking neurons are able to display convincing learning-to-learn behavior, including in the realm of reinforcement learning. Episodic meta-reinforcement learning (Ritter et al., 2018) architectures are also heavily inspired by neuroscientific accounts of complementary learning systems in the brain (McClelland, McNaughton, & O'Reilly, 1995). Both of these examples demonstrate that meta-learning can be used to build more biologically plausible learning algorithms, and thereby highlight that it can act as a bridge between Marr's computational and implementational levels (Marr, 2010).

Finally, the meta-learning perspective not only allows us to connect machine learning and neuroscience via architectural design choices but also via the kinds of tasks that are of interest. Dobs, Martinez, Kell, and Kanwisher (2022), for instance, suggested that functional specialization in neural circuits, which has been widely observed in biological brains, arises as a consequence of task demands. In particular, they found that convolutional neural networks "optimized for both tasks spontaneously segregate themselves into separate systems for faces and objects." Likewise, Yang, Joglekar, Song, Newsome, and Wang (2019) found that training a single recurrent neural network to perform a wide range of cognitive tasks yielded units that were clustered along different functional cognitive processes. Put another way, it seems plausible that functional specialization emerges by training neural networks on multiple tasks. Although this has not been tested so far, we speculate that this also holds in the meta-learning setting, as it involves training on multiple tasks by design. If this were true, we could look at the emerging areas inside a meta-learned model, and use the resulting insights to generate novel predictions about the processes happening in individual brain areas (Kanwisher, Khosla, & Dobs, 2023).

3. Previous research

Meta-learned models are already starting to transform the cognitive sciences today. They allow us to model things that are hard to capture with traditional models such as compositional generalization, language understanding, and model-based reasoning. In this section, we provide an overview of what has been achieved with the help of meta-learning in previous work. We arranged this review into various thematic subcategories. For each of them, we summarize which key findings have been obtained by meta-learning and discuss why these results would have been difficult to obtain using traditional models of learning by appealing to the insights from the previous section.

3.1 Heuristics and cognitive biases

Meta-learning has been previously used to discover algorithms with a limited computational budget that show human-like cognitive biases as we have already alluded to earlier. Dasgupta et al. (2020) trained a neural network on a distribution of probabilistic inference problems while controlling for the number of its hidden units. They found that their model – when restricted to just a single hidden unit – captured many biases in human reasoning, including a conservatism bias and base rate neglect. Likewise, Binz et al. (2022) trained a neural network on a distribution of decision-making problems while controlling for the number of bits needed to represent the network. Their model discovered two previously suggested heuristics in specific environments and made precise prognoses about when these heuristics should be applied. In particular, knowing the correct ranking of features led to one reason decision making, knowing the directions of features led to an equal weighting heuristic, and not knowing about either of them led to strategies that use weighted combinations of features (also see Figs. 4a and 4b).

In both of these studies, meta-learned models offered a novel perspective on results that were previously viewed as contradictory. This was in part possible because meta-learning enabled us to easily manipulate the complexity of the underlying learning algorithm. Although doing so is, at least in theory, also possible within the Bayesian framework, no Bayesian model that captures the full set of findings from Dasgupta et al. (2020) and Binz et al. (2022) has been discovered so far. We hypothesize that this could be because traditional rational process models struggle to capture that human strategy selection is context-dependent even before receiving any direct feedback signal (Mercier & Sperber, 2017). The meta-learned models of Dasgupta et al. (2020) and Binz et al. (2022), on the other hand, were able to readily show context-specific biases when trained on an appropriate task distribution.

3.2 Language understanding

Meta-learning may also help us to answer questions regarding how people process, understand, and produce language. Whether the inductive biases needed to acquire a language are learned from experience or are inherited is one of these questions (Yang & Piantadosi, 2022). McCoy, Grant, Smolensky, Griffiths, and Linzen (2020) investigated how to equip a model with a set of linguistic inductive biases that are relevant to human cognition. Their solution to this problem builds upon the idea of model-agnostic meta-learning (Finn et al., 2017). In particular, they meta-learned the initial weights of a neural network such that the network can adapt itself quickly to new languages using standard gradient-based learning. When being trained on a distribution over languages, these initial weights can be interpreted as universal factors that are shared across all languages. They showed that this approach identifies inductive biases (e.g., a bias for treating certain phonemes as vowels) that are useful for acquiring a language's syllable structure. Although their current work makes limited claims about human language acquisition, their approach be used in future studies to disentangle which inductive biases are learned from experience and which ones are inherited. They additionally argued that a Bayesian modeling approach would only be able to consider a restrictive set of inductive biases as it needs to commit to a particular representation and inference algorithm. In contrast, the meta-learning framework made it easy to implement the intended inductive biases by simply manipulating the distribution of encountered languages.

The ability to compose simple elements into complex entities is at the heart of human language. The property of languages to "make infinite use of finite means" (Chomsky, 2014) is what allows us to make strong generalizations from limited data. For

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

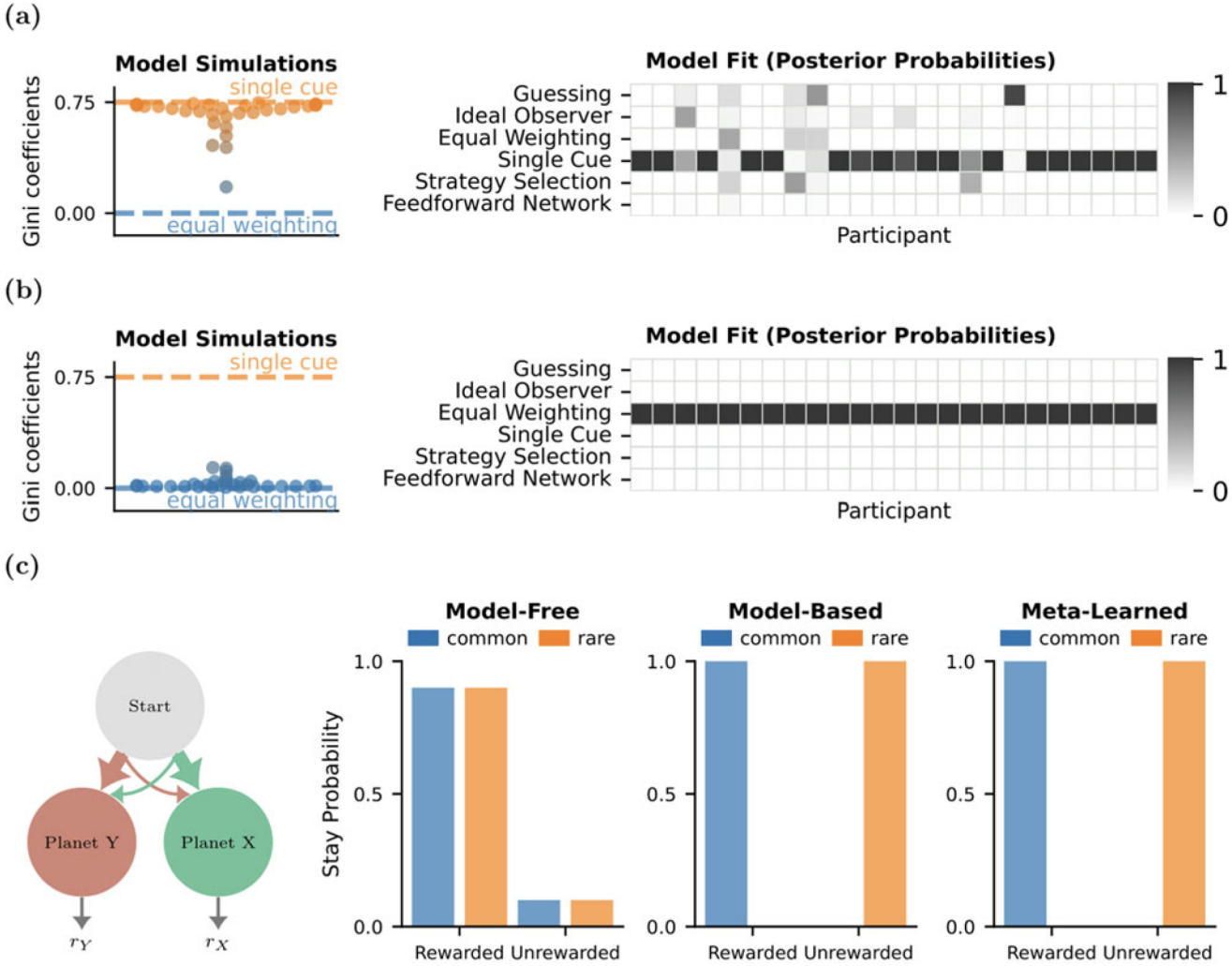

Figure 4. Example results obtained using meta-learned models. (a) In a paired comparison task, a meta-learned model identified a single-cue heuristic as the resource-rational solution when information about the feature ranking was available. Follow-up experiments revealed that people indeed apply this heuristic under the given circumstances. (b) If information about feature directions was available, the same meta-learned model identified an equal weighting heuristic as the resource-rational solution. People also applied this heuristic in the given context (Binz et al., 2022). (c) Wang et al. (2016) showed that meta-learned models can exhibit model-based learning characteristics in the two-step task (Daw et al., 2011) even when they were purely trained through model-free approaches. The plots on the right illustrate the probability of repeating the previous action for different agents (model-free, model-based, meta-learned) after a common or uncommon transition and after a received or omitted reward.

example, people readily understand what it means to "dax twice" or to "dax slowly" after learning about the meaning of the verb "dax." How to build models with a similar proficiency, however, remains an open research question. Lake (2019) showed that a transformer-like neural network can be trained to make such compositional generalizations through meta-learning. Importantly, during meta-learning, his models were adapted to problems that required compositional generalization, and could thereby acquire the skills needed to solve entirely new problems.

Although Lake (2019) argued that meta-learning "has implications for understanding how people generalize compositionally," he did not conduct a direct comparison to human behavior. In a follow-up study, Lake and Baroni (2023) addressed this shortcoming and found that meta-learned models "mimic human systematic generalization in a head-to-head comparison." These results are further corroborated by a recent paper of Jagadish, Binz, Saanum, Wang, and Schulz (2023) which demonstrated that meta-learned models capture human zero-shot compositional inferences in a reinforcement learning setting. However, there also remain open challenges in this context. For example, meta-learned models do not always generalize systematically to longer sequences than those in the training data (Lake, 2019; Lake & Baroni, 2023). How to resolve this issue will be an important challenge for future work.

3.3 Inductive biases

Human cognition comes with many useful inductive biases beyond the ability to reason compositionally. The preference for simplicity is one of these biases (Chater & Vitányi, 2003; Feldman, 2016). We readily extract abstract low-dimensional rules that allow us to generalize entirely new situations. Meta-learning is an ideal tool to build models with similar preferences because we can easily generate tasks based on simple rules and use them for meta-learning, thereby enabling an agent to acquire the desired inductive bias from data.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Toward this end, Kumar, Dasgupta, Cohen, Daw, and Griffiths (2020b) tested humans and meta-reinforcement agents on a grid-based task. People, as well as agents, encountered a series of 7×7 grids. Initially, all tiles were white, but clicking on them revealed their identity as either red or blue. The goal was to reveal all the red tiles while revealing as few blue tiles as possible. There was an underlying pattern that determined how the red tiles were placed, which was either specified by a structured grammar or by a nonstructured process with matched statistics. Humans found it easier to learn in structured tasks, confirming that they have strong priors toward simple abstract rules (Schulz, Tenenbaum, Duvenaud, Speekenbrink, & Gershman, 2017). However, their analysis also indicated that meta-learning is easier on nonstructured tasks than on structured tasks. In follow-up work, they found that this result also holds for agents that were trained purely on the structured version of their task but evaluated on both versions (Kumar et al., 2022a) – a quite astonishing finding considering that one would expect an agent to perform better on the task distribution it was trained on. The authors addressed this mismatch between humans and meta-learned agents by guiding agents during training to reproduce natural language descriptions that people provided to describe a given task. They found that grounding meta-learned agents in natural language descriptions not only improved their performance but also led to more human-like inductive biases, demonstrating that natural language can serve as a source for abstractions within human cognition.

Their line of work uses another interesting technique for training meta-learning agents (Kumar et al., 2022a, 2022b). It does not rely on a hand-designed task distribution but instead involves sampling tasks from the prior distribution of human participants using a technique known as Gibbs sampling with people (Harrison et al., 2020; Sanborn & Griffiths, 2007). Although doing so provides them with a data-set of tasks, no expression of the corresponding prior distribution over them is accessible and, hence, it is nontrivial to define a Bayesian model for the given setting. A meta-learned agent, on the other hand, was readily obtained by training on the collected samples.

3.4 Model-based reasoning

Many realistic scenarios afford two distinct types of learning: model-free and model-based. Model-free learning algorithms directly adjust their strategies using observed outcomes. Model-based learning algorithms, on the other hand, learn about the transition and reward probabilities of an environment, which are then used for downstream reasoning tasks. People are generally thought to be able to perform model-based learning, at least to some extent, and assuming that the problem at hand calls for it (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Kool, Cushman, & Gershman, 2016). Wang et al. (2016) showed that a meta-learned algorithm can display model-based behavior, even if it was trained through a pure model-free reinforcement learning algorithm (see Fig. 4c).

Having a model of the world also acts as the basis for causal reasoning. Traditionally, making causal inferences relies on the notion of Pearl's do-calculus (Pearl, 2009). Dasgupta et al. (2019), however, showed that meta-learning can be used to create models that draw causal inferences from observational data, select informative interventions, and make counterfactual predictions. Although they have not related their model to human data directly, it could in future work serve as the basis to study how people make causal judgments in complex domains and explain why and when they deviate from normative causal theories (Bramley, Dayan, Griffiths, & Lagnado, 2017; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2021).

Together, these two examples highlight that model-based reasoning capabilities can emerge internally in a meta-learned model if they are beneficial for solving the encountered problem. Although there are already many traditional models that can perform such tasks, these models are often slow at run-time as they typically involve Bayesian inference, planning, or both. Meta-learning, on the other hand, "shifts most of the compute burden from inference time to training time [which] is advantageous when training time is ample but fast answers are needed at run-time" (Dasgupta et al., 2019), and may therefore explain how people can perform such intricate computations within a reasonable time frame.

Although model-based reasoning is an emerging property of meta-learned models, it may also be integrated explicitly into such models should it be desired. Jensen, Hennequin, and Mattar (2023) have taken this route, and augmented a standard meta-reinforcement learning agent with the ability to perform temporally extended planning using imagined rollouts. In each time-step, their agent can decide to perform a planning operation instead of directly interacting with the environment (in this case, a spatial navigation task). Their meta-learned agents opted to perform this planning operation consistently after training. Importantly, the model showed striking similarities to patterns of human deliberation by performing more planning early on and with an increased distance to the goal. Furthermore, they found that patterns of hippocampal replays resembled the rollouts of their model.

3.5 Exploration

People do not only have to integrate observed information into their existing knowledge, but they also have to actively determine what information to sample. They constantly face situations that require them to decide whether they should explore something new or whether they should rather exploit what they already know. Previous research suggests that people solve this exploration–exploitation dilemma using a combination of directed and random exploration strategies (Gershman, 2018; Schulz & Gershman, 2019; Wilson, Geana, White, Ludvig, & Cohen, 2014; Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018). Why do people use these particular strategies and not others? Binz and Schulz (2022a) hypothesized that they do so because human exploration follows resource-rational principles. To test this claim, they devised a family of resource-rational reinforcement learning algorithms by combining ideas from meta-learning and information theory. Their meta-learned model discovered a diverse set of exploration strategies, including random and directed exploration, that captured human exploration better than alternative approaches. In this domain, meta-learning offered a direct path toward investigating the hypothesis that people try to explore optimally but are subject to limited computational resources, whereas designing hand-crafted models for studying the same question would have been more intricate.

It is not only important to decide how to explore, but also to decide whether exploration is worthwhile in the first place. Lange and Sprekeler (2020) studied this question using the meta-learning framework. Their meta-learned agents are able to flexibly interpolate between implementing exploratory learning

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

behaviors and hard-coded, nonlearning strategies. Importantly, which behavior was realized crucially depended on environmental properties, such as the diversity of the task distribution, the task complexity, and the agent's lifetime. They showed, for instance, that agents with a short lifetime should opt for small rewards that are easy to find, whereas agents with an extended lifetime should spend their time exploring the environment. The study of Lange and Sprekeler (2020) clearly demonstrates that meta-learning makes it conceptually easy to iterate over different environmental assumptions inside a rational analysis of cognition. They only had to modify the environment as desired, followed by rerunning their meta-learning procedure. In contrast, traditional modeling approaches would require hand-designing a new optimal agent each time an environmental change occurs.

3.6 Cognitive control

Humans are remarkable at adapting to task-specific demands. The processes behind this ability are collectively referred to as cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Cohen (2017) even argues that "the capacity for cognitive control is perhaps the most distinguishing characteristic of human behavior." It should therefore come as no surprise that cognitive control has received a significant amount of attention from a computational perspective (Botvinick & Cohen, 2014; Collins & Frank, 2013). Recently, some of these computational investigations have been extended to the meta-learning framework.

The ability to adjust computational resources as needed is one hallmark of cognitive control. Moskovitz, Miller, Sahani, and Botvinick (2022) proposed a meta-learned model with such characteristics. Their model learns a simple default policy – similar to the model of Binz and Schulz (2022a) – that can be overwritten by a more complex one if necessary. They demonstrate that this model is not only able to capture behavioral phenomena from the cognitive control literature but also known effects in decision-making and reinforcement learning tasks, thereby linking the three domains. Importantly, their study highlights that the meta-learning framework offers the means to account for multiple computational costs instead of just a single one – in this case, a cost for implementing the default policy and one for deviating from it.

Taking contextual cues into consideration is another vital aspect of cognitive control. Dubey, Grant, Luo, Narasimhan, and Griffiths (2020) implemented this idea in the meta-learning framework. In their model, contextual cues determine the initialization of a task-specific neural network that is then trained using model-agnostic meta-learning. They showed that such a model captures "the context-sensitivity of human behavior in a simple but well-studied cognitive control task." Furthermore, they demonstrated that it scales well to more complex domains (including tasks from the MuJoCo [Todorov, Erez, & Tassa, 2012], CelebA [Liu, Luo, Wang, & Tang, 2015], and MetaWorld [Yu et al., 2020] benchmarks), thereby opening up new opportunities for modeling human behavior in naturalistic scenarios.

4. Why is not everything meta-learned?

We have laid out different arguments that make meta-learning a useful tool for constructing cognitive models, but it is important to note that we do not claim that meta-learning is the ultimate solution to every modeling problem. Instead, it is essential to understand when meta-learning is the right tool for the job and when not.

4.1 Lack of interpretability

Perhaps its most significant detriment is that meta-learning never provides us with analytical solutions that we can inspect, analyze, and reason about. In contrast to this, some Bayesian models have analytical solutions. Take as an example the data-generating distribution that we encountered earlier (Equations (1)–(2)). For these assumptions, a closed-form expression of the posterior predictive distribution is available. By looking at this closed-form expression, researchers have generated new predictions and subsequently tested them empirically (Daw et al., 2008; Dayan & Kakade, 2000; Gershman, 2015). Performing the same kind of analysis with a meta-learned model is not as straightforward. We do not have access to an underlying mathematical expression, which makes a structured exploration of theories much harder.

That being said, there are still ways to analyze a meta-learned model's behavior. For one, it is possible to use model architectures that facilitate interpretability. Binz et al. (2022) relied on this approach and designed a neural network architecture that produced weights of a probit regression model that were then used to cluster applied strategies into different categories. Doing so enabled them to identify which strategy was used by their meta-learned model in a particular situation.

Recently, researchers have also started to use tools from cognitive psychology to analyze the behavior of black-box models (Bowers et al., 2022; Rich & Gureckis, 2019; Ritter, Barrett, Santoro, & Botvinick, 2017; Schulz & Dayan, 2020). For example, it is possible to treat such models just like participants in a psychological experiment and use the collected data to analyze their behavior similar to how psychologists would analyze human behavior (Binz & Schulz, 2023; Dasgupta et al., 2022; Rahwan et al., 2019; Schramowski, Turan, Andersen, Rothkopf, & Kersting, 2022). We believe that this approach has great potential for analyzing increasingly capable and opaque artificial agents, including those obtained via meta-learning.

4.2 Intricate training processes

When using the meta-learning framework, one also has to deal with the fact that training neural networks is complex and takes time. Neural network models contain many moving parts, like weight initializations or the used optimizer, that have to be chosen appropriately such that training can take off in the first place, and training itself may take hours or days until it is finished. When we want to modify assumptions in the data-generating distribution, we have to retrain the whole system from scratch altogether. Thus, although the process of iterating over different environmental assumptions is conceptually straightforward in the meta-learning framework, it may be time consuming. Bayesian models can, in comparison, sometimes be more quickly adapted to changes in environmental assumptions. To illustrate this, let us assume that you wanted to explain human behavior through a meta-learned model that was trained on the data-generating distribution from Equations (1)–(2), but found that the resulting model does not fit the observed data well. Next, you want to consider the alternative hypothesis that people assume a nonstationary environment. Although this modification could be done quickly in the corresponding Bayesian model, the

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

meta-learning framework requires retraining on newly generated data.

There is, furthermore, no guarantee that a fully converged meta-learned model implements a Bayes-optimal learning algorithm. Indeed, there are reported cases in which meta-learning failed to find the Bayes-optimal solution (Wang et al., 2021). In simple scenarios, like our insect length example, we can resolve this issue by comparing to analytical solutions. This kind of reasoning applies to some of the settings in which meta-learning has been used to study human behavior. For example, for the exploration studies discussed in the previous section, it has been shown that meta-learned models closely approximate the (tractable but computationally expensive) Bayes-optimal algorithm (Duan et al., 2016; Wang et al., 2016). However, in more complex scenarios, it is impossible to verify that a meta-learned algorithm is optimal. We believe that this issue can be somewhat mitigated by validating meta-learned models in various ways. For example, we may get an intuition for the correspondence between a meta-learned model and an intractable Bayes-optimal algorithm by comparing to other approximate inference techniques (as done in Binz et al., 2022) or to symbolic models (as done in Lake & Baroni, 2023). In the end, however, we believe that this issue is still an open problem and that future work needs to come up with novel techniques to verify meta-learned models. Nevertheless, this is already a step forward as verifying solutions is often easier than generating them.

4.3 Meta-learned or Bayesian inference?

In summary, both frameworks – meta-learning and Bayesian inference – have their unique strengths and weaknesses. The meta-learning framework does and will not replace Bayesian inference but complement it. It broadens our available toolkit and enables researchers to study questions that were previously out of reach. However, there are certainly situations in which traditional Bayesian inference is a more appropriate modeling choice as we have outlined in this section.

5. The role of neural networks

Most of the points we have discussed so far are agnostic regarding the function approximator implementing the meta-learned algorithm. However, at the same time, we have appealed to neural networks at various points throughout the text. When one looks at prior work, it can also be observed that neural networks are the predominant model class in the meta-learning setting. Why is that the case? In addition to their universality, neural networks offer one big opportunity: They provide a flexible framework for engineering different types of inductive biases into a computational model (Goyal & Bengio, 2022). In the following section, we will highlight three examples of how previous work has accomplished this. For each of these examples, we take a concept from psychology, and show how it can be readily accommodated in a meta-learned model.

Perhaps one of the most persuasive ideas in cognitive modeling is that of gradient-based learning. It is not only at the heart of one of the most influential models – the Rescorla–Wagner model (Gershman, 2015; Rescorla, 1972) – but also features prominently in many other theories of human learning, such as connectionist models (Rumelhart et al., 1988). Even though the earlier outlined meta-learning procedure relies on gradient-based learning in the outer-loop, the resulting inner-loop dynamics must bear no resemblance to gradient descent. However, it is possible to construct meta-learned models whose inner-loop updates rely on gradient-based learning. Finn et al. (2017) proposed a meta-learning technique known as model-agnostic meta-learning that finds optimal initial parameters of a feedforward neural network that is subsequently trained via gradient descent. The idea is that these optimal initial parameters allow the feedforward network to generalize to multiple tasks in a minimal number of gradient steps. Although their general setup is similar to the one we discussed, it leads to models that learn via gradient descent instead of models that implement a learning algorithm inside the dynamics of a recurrent neural network. Kirsch and Schmidhuber (2021) recently brought these two approaches together into a single model. Their proposed architecture consists of multiple recurrent neural networks that interact with each other. Importantly, they showed that one particular configuration of these networks could implement backpropagation in the forward pass, thereby being able to perform gradient-based learning in a memory-based system.

Exemplar-based models – like the generalized category model (Nosofsky, 2011) – are one of the most prominent approaches for modeling how people categorize items into different classes (Kruschke, 1990; Shepard, 1987). They categorize a new instance based on the estimated similarity between that instance and previously seen examples. Recently, meta-learned models with exemplar-based reasoning abilities have been proposed for the task of few-shot classification, in which a classifier must generalize based on a training set containing only a few examples. Matching networks (Vinyals et al., 2016) accomplish this by classifying a new data-point using a similarity-weighted combination of categories in the training set. Importantly, similarity is computed over a learned embedding space, thereby ensuring that the model can scale to high-dimensional stimuli. Follow-up work has taken inspiration from another hugely influential model of human category learning and replaced the exemplar-based mechanism used in matching networks with one based on category prototypes (Snell, Swersky, & Zemel, 2017).

Finally, making inferences using similarities to previous experiences is not only useful for supervised learning but also in the reinforcement learning setting. In the reinforcement learning literature, the ability to store and recollect states or trajectories for later use is studied under the name of episodic memory (Lengyel & Dayan, 2007). It has been argued that episodic memory could be the key to explaining human performance in naturalistic environments (Gershman & Daw, 2017). Episodic memory also plays a crucial role in neuroscience, with studies showing that highly rewarding instances are stored in the hippocampus and made available for recall as and when required (Blundell et al., 2016). Ritter et al. (2018) build upon the neural episodic control idea from Pritzel et al. (2017) and use a differential neural dictionary for episodic recall in the context of meta-learning. Their dictionary stores encodings from previously experienced tasks, which can then be later queried as needed. With this addition, their meta-learned model is able to recall previously discovered policies, retrieve memories using category examples, handle compositional tasks, re-instate memories while traversing the environment, and recover a learning strategy people use in a neuroscience-inspired task.

In summary, human cognition comes with a variety of inductive biases and neural networks provide flexible ways to easily incorporate them into meta-learned models of cognition. We have outlined three such examples in the section, demonstrating

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

how to integrate gradient-based learning, exemplar- and prototype-based reasoning, and episodic memory into a meta-learned model. There are, furthermore, many other inductive biases for neural network architectures that could be used in the context of meta-learning but have not been yet. Examples include networks that perform differentiable planning (Farquhar, Rocktäschel, Igl, & Whiteson, 2017; Tamar, Wu, Thomas, Levine, & Abbeel, 2016), extract object-based representations (Piloto, Weinstein, Battaglia, & Botvinick, 2022; Sancaktar, Blaes, & Martius, 2022), or modify their own connections through synaptic plasticity (Miconi, Rawal, Clune, & Stanley, 2020; Schlag, Irie, & Schmidhuber, 2021).

6. Toward a domain-general model of human learning

What does the future hold for meta-learning? The current generation of meta-learned models of cognition is almost exclusively trained on the data-generating distribution of a specific problem family. Although this training process enables them to generalize to new tasks inside this problem family, they are unlikely to generalize to completely different domains. We would, for example, not expect a meta-learned algorithm to perform a challenging maze navigation task if it was only trained to predict the lengths of insect species.

Although domain-specific models have (and will continue to) provide answers to important research questions, we agree with Newell (1992) that "unified theories of cognition are the only way to bring this wonderful, increasing fund of knowledge under intellectual control." Ideally, such a unified theory should manifest itself in a domain-general cognitive model that cannot only solve prediction tasks but is also capable of human-like decision making (Gigerenzer & Gaissmaier, 2011), category learning (Ashby et al., 2005), navigation (Montello, 2005), problem-solving (Newell et al., 1972), and so forth. We consider the meta-learning framework the ideal tool for accomplishing this goal as it allows us to compile arbitrary assumptions about an agent's beliefs of the world (arguments 1 and 2) and its computational architecture (arguments 3 and 4) into a cognitive model.

To obtain such a domain-general cognitive model via meta-learning, however, a few challenges need to be tackled. First of all, there is the looming question of how a data-generating distribution that contains many different problems should be constructed. Here, we may take inspiration from the machine learning community, where researchers have devised generalist agents by training neural networks on a large set of problems (Reed et al., 2022). Team et al. (2023) have recently shown that this is a promising path for scaling up meta-learning models. They trained a meta-reinforcement learning agent on a vast open-ended world with over 10^{40} possible tasks. The resulting agent can adapt to held-out problems as quickly as humans, and "displays on-the-fly hypothesis-driven exploration, efficient exploitation of acquired knowledge, and can successfully be prompted with first-person demonstrations." In the same vein, we may come up with a large collection of tasks that are more commonly used to study human behavior (Miconi, 2023; Molano-Mazon et al., 2022; Yang et al., 2019), and use them to train a meta-learned model of cognition.

Language will likely play an important role in future meta-learning systems. We do not want a system that learns every task from scratch via trial and error but one that can be provided with a set of instructions similar to how a human subject would be instructed in a psychological experiment. Having agents capable of language will not only enable them to understand new tasks in a zero-shot manner but may also facilitate their cognitive abilities. It, for example, allows them to decompose tasks into subtasks, learn from other agents, or generate explanations (Colas, Karch, Moulin-Frier, & Oudeyer, 2022). Fortunately, our current best language models (Brown et al., 2020; Chowdhery et al., 2022) and meta-learning systems are both based on neural networks. Thus, integrating language capabilities into a meta-learned model of cognition should – at least conceptually – be fairly straightforward. Doing so would furthermore enable such models to harvest the compositional nature of language to make strong generalizations to tasks outside of the meta-learning distribution. The potential for this was highlighted in a recent study (Riveland & Pouget, 2022) which found that language-conditioned recurrent neural network models can perform entirely novel psychophysical tasks with high accuracy.

Moreover, a sufficiently general model of human cognition must not only be able to select among several given options, like in a decision-making or category learning setting, but it also needs to maneuver within a complex world. For this, it needs to perceive and process high-dimensional visual stimuli, it needs to control a body with many degrees of freedom, and it needs to actively engage with other agents. Many of these problems have been at the heart of the deep learning community (Hill et al., 2020; McClelland, Hill, Rudolph, Baldridge, & Schütze, 2020; Strouse, McKee, Botvinick, Hughes, & Everett, 2021; Team et al., 2021), and it will be interesting to see whether the solutions developed there can be integrated into a meta-learned model of cognition.

Finally, there are also some challenges on the algorithmic side that need to be taken into account. In particular, it is unclear how far currently used model architectures and outer-loop learning algorithms scale. Although contemporary meta-learning algorithms are able to find approximately Bayes-optimal solutions to simple problems, they sometimes struggle to do so on more complex ones (e.g., as in the earlier discussed work of Wang et al., 2021). Therefore, it seems likely that simply increasing the complexity of the meta-learning distribution will not be sufficient – we will also need model architectures and outer-loop learning algorithms that can handle increasingly complex data-generating distributions. The transformer architecture (Vaswani et al., 2017), which has been very successful at training large language models (Brown et al., 2020; Chowdhery et al., 2022), provides one promising candidate, but there could be countless other (so far undiscovered) alternatives.

Thus, taken together, there are still substantial challenges involved in creating a domain-general meta-learned model of cognition. In particular, we have argued in this section that we need to (1) meta-learn on more diverse task distributions, (2) develop agents that can parse instructions in the form of natural language, (3) embody these agents in realistic problem settings, and (4) find model architectures that scale to these complex problems. Figure 5 summarizes these points graphically.

7. Conclusion

Most computational models of human learning are hand-designed, meaning that at some point a researcher sat down and defined how they behave. Meta-learning starts with an entirely different premise. Instead of designing learning algorithms by hand, one trains a system to achieve its goals by repeatedly letting it interact with an environment. Although this seems

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

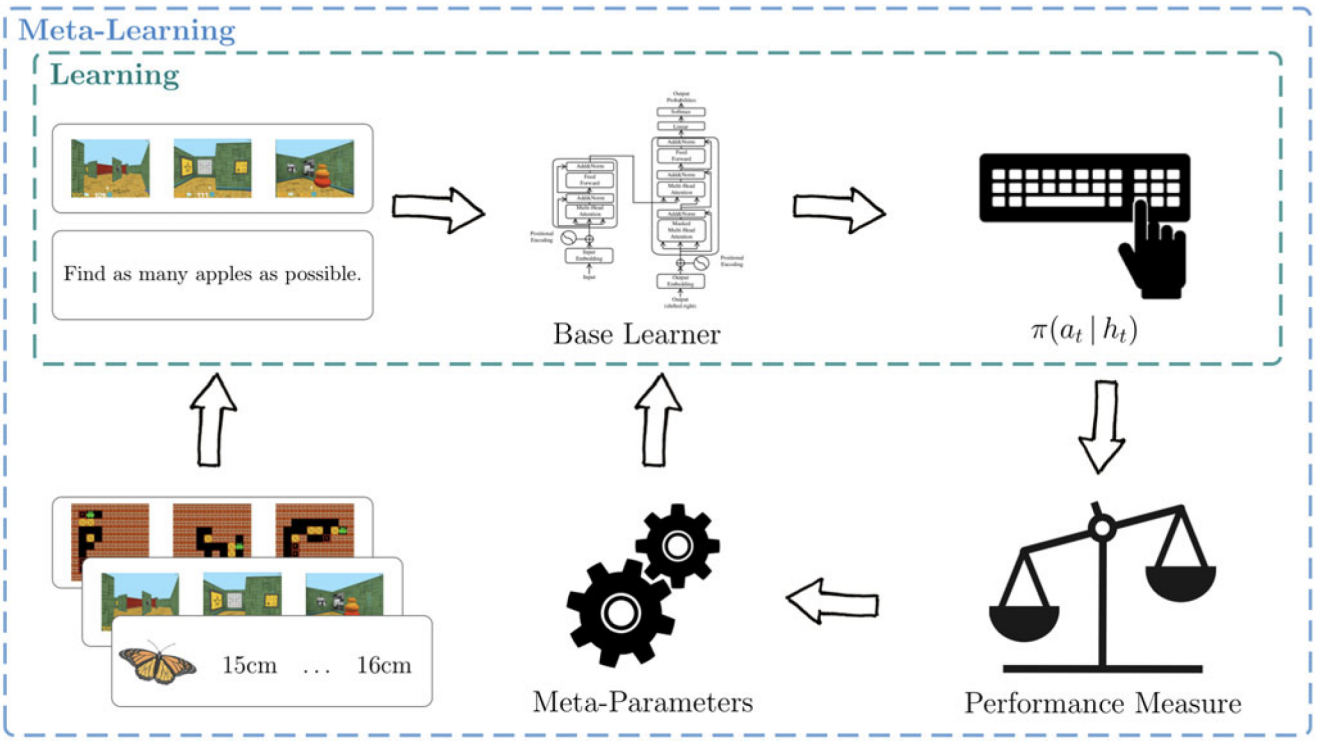

Figure 5. Illustration of how a domain-general meta-learned model of cognition could look like. Modifications include training on more diverse task distributions, providing natural language instructions as additional inputs, and relying on scalable model architectures.

quite different from traditional models of learning on the surface, we have highlighted that the meta-learning framework actually has a deep connection to the idea of Bayesian inference, and thereby to the rational analysis of cognition. Using this connection as a starting point, we have highlighted several advantages of the meta-learning framework for constructing rational models of cognition. Together, our arguments demonstrate that meta-learning cannot only be applied in situations where Bayesian inference is impossible but also facilitates the inclusion of computational constraints and neuroscientific insights into rational models of human cognition. Earlier criticisms of the rational analysis of cognition have repeatedly pointed out that "rational Bayesian models are significantly unconstrained" and that they should be "developed in conjunction with mechanistic considerations to offer substantive explanations of cognition" (Jones & Love, 2011). Likewise, Bowers and Davis (2012) argued that to understand human cognition "important constraints [must] come from biological, evolutionary, and processing (algorithmic) considerations." We believe that the meta-learning framework provides the ideal opportunity to address these criticisms as it allows for a painless integration of flexible algorithmic (often biologically inspired) mechanisms.

It is worth pointing out that meta-learning can be also motivated by taking neural networks as a starting point. From this perspective, it bridges two of the most popular theories of cognition – Bayesian models and connectionism – by bringing the scalability of neural network models into the rational analysis of cognition. The blending of Bayesian models and neural networks situates the meta-learning framework at the heart of the debate on whether cognition is best explained by emergentist or structured probabilistic approaches (Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; McClelland et al., 2010). Like traditional connectionist approaches, meta-learning provides a means to explain how cognition could emerge as a system repeatedly interacts with an environment. Whether the current techniques used for meta-learning mirror the emergence of cognitive processes in people however remains an open question. Personally, we believe that this is unlikely and that there are more elaborate processes in play during human meta-learning than the gradient descent-based algorithms that are commonly used for training neural networks (Schulze Buschoff, Schulz, & Binz, 2023). To study this question systemically, we would need to look at human behavior across much longer timescales (e.g., developmental or evolutionary). Yet, at the same time, meta-learning does not limit itself to a purely emergentist perspective. The modern neural network toolbox allows researchers to flexibly integrate additional structure and inductive biases into a model by adjusting the underlying network architecture – as we have argued in Section 5 – thereby preserving a key advantage of structured probabilistic approaches. How much hand-crafting within the network architecture is needed ultimately depends on the designer's goals. The meta-learning framework is agnostic to this and allows it to range from almost nothing to a substantial amount.

We believe that meta-learning provides a powerful tool to scale up psychological theories to more complex settings. However, at the same time, meta-learning has not delivered on this promise yet. Existing meta-learned models of cognition are typically applied to classical scenarios where established models already exist. Thus, we have to ask: What prevents the application to more complex and general paradigms? First, such paradigms themselves have to be developed. Fortunately, there is currently a trend toward measuring human behavior on more naturalistic tasks (Brändle, Binz, & Schulz, 2022a; Brändle, Stocks,

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Tenenbaum, Gershman, & Schulz, 2022b; Schulz et al., 2019), and it will be interesting to see what role meta-learning will play in modeling behavior in such settings. Furthermore, meta-learning can be intricate and time consuming. We hope that the present article – together with the accompanying code examples – makes this technique less opaque and more accessible to a wider audience. Future advances in hardware will likely make the meta-learning process quicker and we are therefore hopeful that meta-learning can ultimately fulfill its promise of identifying plausible models of human cognition in situations that are out of reach for hand-designed algorithms.

Acknowledgments. The authors thank Sreejan Kumar, Tobias Ludwig, Dominik Endres, and Adam Santoro for their valuable feedback on an earlier draft.

Financial support. This work was funded by the Max Planck Society, the Volkswagen Foundation, as well as the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC2064/1-390727645.

Competing interest. None.

Notes

1. Based on our earlier definition, it is at this point strictly speaking not a learning algorithm at all as it does not improve with additional data.
2. https://github.com/marcelbinz/meta-learned-models.
3. There has been a long-standing conceptual debate in cognitive psychology on whether to view Bayesian models as normative standards or descriptive tools. We believe that this debate is beyond the scope of the current article and thus refer the reader to earlier work for an in-depth discussion (Bowers & Davis, 2012; Griffiths, Chater, Norris, & Pouget, 2012; Jones & Love, 2011; Tauber, Navarro, Perfors, & Steyvers, 2017; Zednik & Jäkel, 2016). We only want to add that the framework outlined here is agnostic to this issue – meta-learned models may serve as both normative standards and descriptive tools.
4. In principle, one could select arbitrarily flexible functional forms, such as mixtures of normal distributions or discretized distributions with small bin sizes, which would reduce the accompanying approximation error.
5. This only holds for standard variational inference but not for more advanced methods that involve amortization such as variational autoencoders (Kingma & Welling, 2013).
6. Note that although it is possible to apply some Bayesian models (e.g., non-parametric methods) in this setting, we would have to contend with making arbitrary assumptions about the likelihood function, causing a loss of optimality guarantees.
7. Having said that, it is possible to approximate it under certain circumstances and different authors have applied such approximations to study both human and animal cognition (Chater & Vitányi, 2003; Gauvrit, Zenil, Delahaye, & Soler-Toscano, 2014; Gauvrit, Zenil, & Tegnér, 2017; Griffiths, Daniels, Austerweil, & Tenenbaum, 2018; Zenil, Marshall, & Tegnér, 2015).

References

Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, *62*(3), 547–554.
Anderson, J. R. (2013a). *The adaptive character of thought*. Psychology Press.
Anderson, J. R. (2013b). *The architecture of cognition*. Psychology Press.
Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*(1), 149–178.
Bates, C. J., & Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological Review*, *127*(5), 891.
Baxter, J. (1998). Theoretical models of learning to learn. In S. Thrun & L. Pratt (Eds.), *Learning to learn* (pp. 71–94). Springer.
Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., & Maass, W. (2018). Long short-term memory and learning-to-learn in networks of spiking neurons. *Advances in Neural Information Processing Systems*, *31*, 795–805.
Bengio, Y., Bengio, S., & Cloutier, J. (1991). Learning a synaptic learning rule. In *IJCNN-91-Seattle International Joint Conference on Neural Networks, Seattle, WA, USA* (Vol. 2, p. 969).
Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In B. Bernheim, S. DellaVigna, & D. Laibson (Eds.), *Handbook of behavioral economics: Applications and foundations* (Vol. 2, pp. 69–186). North-Holland.
Binmore, K. (2007). Rational decisions in large worlds. *Annales d'Economie et de Statistique*, *No. 86*, 25–41.
Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2022). Heuristics from bounded meta-learned inference. *Psychological Review*, *129*(5), 1042–1077.
Binz, M., & Schulz, E. (2022a). Modeling human exploration through resource-rational reinforcement learning. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.), *Advances in neural information processing systems* (pp. 31755–31768). Curran Associates, Inc. https://openreview.net/forum?id=W1MUJv5zaXP.
Binz, M., & Schulz, E. (2022b). Reconstructing the Einstellung effect. *Computational Brain & Behavior*, *6*, 1–17.
Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(6), e2218523120.
Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
Blundell, C., Uria, B., Pritzel, A., Li, Y., Ruderman, A., Leibo, J. Z., … Hassabis, D. (2016). Model-free episodic control. *arXiv preprint arXiv:1606.04460*.
Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624.
Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science*, *38*(6), 1249–1285.
Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389.
Bowers, J. S., Malhotra, G., Dujmović, M., Montero, M. L., Tsvetkov, C., Biscione, V., … Blything, R. (2022). Deep problems with neural network models of human vision. *Behavioral and Brain Sciences*, *46*, 1–74.
Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301.
Brändle, F., Binz, M., & Schulz, E. (2022a). Exploration beyond bandits. In I. Cogliati Dezza, E. Schulz, & C. M. Wu (Eds.), *The drive for knowledge: The science of human information seeking* (pp. 147–168). Cambridge University Press. doi:10.1017/9781009026949.008
Brändle, F., Stocks, L. J., Tenenbaum, J. B., Gershman, S. J., & Schulz, E. (2022b). Intrinsically motivated exploration as empowerment. *PsyArXiv*. January 14.
Brighton, H., & Gigerenzer, G. (2012). Are rational actor models "rational" outside small worlds. In S. Okasha & K. Binmore (Eds.), *Evolution and rationality: Decisions, co-operation, and strategic behavior* (pp. 84–109). Cambridge University Press.
Bromberg-Martin, E. S., Matsumoto, M., Hong, S., & Hikosaka, O. (2010). A pallidus–habenula–dopamine pathway signals inferred stimulus values. *Journal of Neurophysiology*, *104*(2), 1068–1076.
Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.
Chaitin, G. J. (1969). On the simplicity and speed of programs for computing infinite sets of natural numbers. *Journal of the ACM (JACM)*, *16*(3), 407–422.
Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65.
Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.
Chomsky, N. (2014). *Aspects of the theory of syntax* (Vol. 11). MIT Press.
Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., … Fiedel, N. (2022). Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.
Cohen, J. D. (2017). Cognitive control: Core constructs and current considerations. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 1–28). Wiley Blackwell.
Colas, C., Karch, T., Moulin-Frier, C., & Oudeyer, P.-Y. (2022). Language and culture internalization for human-like autotelic AI. *Nature Machine Intelligence*, *4*(12), 1068–1076.
Collins, A. G., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190.
Corner, A., & Hahn, U. (2013). Normative theories of argumentation: Are some norms better than others? *Synthese*, *190*(16), 3579–3610.
Courville, A. C., & Daw, N. D. (2008). The rat as particle filter. In J. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (pp. 369–376). Curran Associates, Inc.
Cover, T. M. (1999). *Elements of information theory*. John Wiley.
Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(48), 30055–30062.
Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics. In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 97–118). Oxford University Press.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Dasgupta, I., & Gershman, S. J. (2021). Memory as a computational resource. *Trends in Cognitive Sciences*, *25*(3), 240–251.

Dasgupta, I., Lampinen, A. K., Chan, S. C., Creswell, A., Kumaran, D., McClelland, J. L., & Hill, F. (2022). Language models show human-like content effects on reasoning. *arXiv preprint arXiv:2207.07051*.

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25.

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review*, *127*(3), 412.

Dasgupta, I., Wang, J., Chiappa, S., Mitrovic, J., Ortega, P., Raposo, D., … Kurth-Nelson, Z. (2019). Causal reasoning from meta-reinforcement learning. *arXiv preprint arXiv:1901.08162*.

Daw, N. D., Courville, A. C., & Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind* (pp. 431–452). Oxford Academic.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

Dayan, P., & Kakade, S. (2000). Explaining away in weight space. *Advances in Neural Information Processing Systems*, *13*, 430–436.

Dobs, K., Martinez, J., Kell, A. J., & Kanwisher, N. (2022). Brain-like functional specialization emerges spontaneously in deep neural networks. *Science Advances*, *8*(11), eabl8913.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*(4–6), 495–506.

Duan, Y., Schulman, J., Chen, X., Bartlett, P. L., Sutskever, I., & Abbeel, P. (2016). RL2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*.

Dubey, R., Grant, E., Luo, M., Narasimhan, K., & Griffiths, T. (2020). Connecting context-specific adaptation in humans to meta-learning. *arXiv preprint arXiv:2011.13782*.

Duff, M. O. (2003). *Optimal learning: Computational procedures for Bayes-adaptive Markov decision processes* [Unpublished PhD thesis]. University of Massachusetts Amherst.

Farquhar, G., Rocktäschel, T., Igl, M., & Whiteson, S. (2017). TreeQN and ATreeC: Differentiable tree-structured models for deep reinforcement learning. *arXiv preprint arXiv:1710.11417*.

Feldman, J. (2016). The simplicity principle in perception and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(5), 330–340.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated machine learning* (pp. 3–33). Springer.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126–1135).

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Gauvrit, N., Zenil, H., Delahaye, J.-P., & Soler-Toscano, F. (2014). Algorithmic complexity for short binary strings applied to psychology: A primer. *Behavior Research Methods*, *46*(3), 732–744.

Gauvrit, N., Zenil, H., & Tegnér, J. (2017). The information-theoretic and algorithmic approach to human, animal, and artificial cognition. In G. Dodig-Crnkovic & R. Giovagnoli (Eds.), *Representation and reality in humans, other living organisms and intelligent machines* (pp. 117–139). Springer.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 721–741.

Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLoS Computational Biology*, *11*(11), e1004567.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, *68*, 101.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(5), 936.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B: Methodological*, *41*(2), 148–177.

Goyal, A., & Bengio, Y. (2022). Inductive biases for deep learning of higher-level cognition. *Proceedings of the Royal Society A*, *478*(2266), 20210068.

Grant, E., Finn, C., Levine, S., Darrell, T., & Griffiths, T. (2018). Recasting gradient-based meta-learning as hierarchical Bayes. In *6th international conference on learning representations, ICLR 2018*.

Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., & Lieder, F. (2019). Doing more with less: Meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, *29*, 24–30.

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.

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, *138*(3), 415–422.

Griffiths, T. L., Daniels, D., Austerweil, J. L., & Tenenbaum, J. B. (2018). Subjective randomness as statistical inference. *Cognitive Psychology*, *103*, 85–109.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). Cambridge University Press.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767–773.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*(1), 51.

Harrison, P., Marjieh, R., Adolfi, F., van Rijn, P., Anglada-Tort, M., Tchernichovski, O., … Jacoby, N. (2020). Gibbs sampling with people. *Advances in Neural Information Processing Systems*, *33*, 10659–10671.

Hill, F., Lampinen, A., Schneider, R., Clark, S., Botvinick, M., McClelland, J. L., & Santoro, A. (2020). Environmental drivers of systematicity and generalization in a situated agent. In *International conference on learning representations*. Retrieved from https://openreview.net/forum?id=SklGryBtwr

Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the 6th annual conference on computational learning theory* (pp. 5–13).

Hinton, G. E., & Zemel, R. (1993). Autoencoders, minimum description length and Helmholtz free energy. *Advances in Neural Information Processing Systems*, *6*, 3–10.

Hochreiter, S., Younger, A. S., & Conwell, P. R. (2001). Learning to learn using gradient descent. In *International conference on artificial neural networks* (pp. 87–94).

Hoppe, D., & Rothkopf, C. A. (2016). Learning rational temporal eye movement strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8332–8337.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Jagadish, A. K., Binz, M., Saanum, T., Wang, J. X., & Schulz, E. (2023). Zero-shot compositional reinforcement learning in humans. https://doi.org/10.31234/osf.io/ymve5.

Jensen, K. T., Hennequin, G., & Mattar, M. G. (2023). A recurrent network model of planning explains hippocampal replay and human behavior. *bioRxiv*, 2023-01.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*(2), 183–233.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237.

Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask "why" questions of minds and brains. *Trends in Neurosciences*, *46*(3), 240–254.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*.

Kirsch, L., & Schmidhuber, J. (2021). Meta learning backpropagation and improving it. *Advances in Neural Information Processing Systems*, *34*, 14122–14134.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge University Press.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, *1*(1), 1–7.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control payoff? *PLoS Computational Biology*, *12*(8), e1005090.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Kruschke, J. (1990). Alcove: A connectionist model of human category learning. *Advances in Neural Information Processing Systems*, *3*, 649–655.

Kumar, S., Correa, C. G., Dasgupta, I., Marjieh, R., Hu, M., Hawkins, R. D., … Griffiths, T. L. (2022a). Using natural language and program abstractions to instill human inductive biases in machines. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.), *Advances in neural information processing systems* (pp. 167–180). Curran Associates, Inc. https://openreview.net/forum?id=buXZ7nIqiwE.

Kumar, S., Dasgupta, I., Cohen, J., Daw, N., & Griffiths, T. (2020b). Meta-learning of structured task distributions in humans and machines. In *International conference on learning representations*.

Kumar, S., Dasgupta, I., Cohen, J. D., Daw, N. D., & Griffiths, T. L. (2020a). Meta-learning of structured task distributions in humans and machines. *arXiv preprint arXiv:2010.02317*.

Kumar, S., Dasgupta, I., Marjieh, R., Daw, N. D., Cohen, J. D., & Griffiths, T. L. (2022b). Disentangling abstraction from statistical pattern matching in human and machine learning. *arXiv preprint arXiv:2204.01437*.

Lake, B. M. (2019). Compositional generalization through meta sequence-to-sequence learning. *Advances in Neural Information Processing Systems*, *32*, 9791–9801.

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*, 1–7.

Lange, R. T., & Sprekeler, H. (2020). Learning not to learn: Nature versus nurture in silico. *arXiv preprint arXiv:2010.04466*.

Lengyel, M., & Dayan, P. (2007). Hippocampal contributions to control: The third way. *Advances in Neural Information Processing Systems*, *20*, 889–896.

Lewis, D. (1999). Why conditionalize? In Lewis D. (Ed.) , *Papers in metaphysics and epistemology* (Vol. 2, pp. 403–407). Cambridge University Press. doi:10.1017/CBO9780511625343.024

Li, Z., Zhou, F., Chen, F., & Li, H. (2017). Meta-SGD: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*.

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological Review*, *124*(6), 762.

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015, December). Deep learning face attributes in the wild. In *Proceedings of international conference on computer vision (ICCV)*.

Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review*, *22*(5), 1193–1215.

Lueckmann, J.-M., Boelts, J., Greenberg, D., Goncalves, P., & Macke, J. (2021). Benchmarking simulation-based inference. In *International conference on artificial intelligence and statistics* (pp. 343–351).

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356.

McClelland, J. L., Hill, F., Rudolph, M., Baldridge, J., & Schütze, H. (2020). Placing language in an integrated understanding system: Next steps toward human-level performance in neural language models. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(42), 25966–25974.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419.

McCoy, R. T., Grant, E., Smolensky, P., Griffiths, T. L., & Linzen, T. (2020). Universal linguistic inductive biases via meta-learning. *arXiv preprint arXiv:2006.16324*.

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard University Press.

Miconi, T. (2023). A large parametrized space of meta-reinforcement learning tasks. *arXiv preprint arXiv:2302.05583*.

Miconi, T., Rawal, A., Clune, J., & Stanley, K. O. (2020). Backpropamine: Training self-modifying neural networks with differentiable neuromodulated plasticity. *arXiv preprint arXiv:2002.10585*.

Mikulik, V., Delétang, G., McGrath, T., Genewein, T., Martic, M., Legg, S., & Ortega, P. (2020). Meta-trained agents implement Bayes-optimal agents. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 18691–18703). Curran. Retrieved from https://proceedings.neurips.cc/paper/2020/file/d902c3ce47124c66ce615d5ad9ba304f-Paper.pdf

Mitchell, T. M. (1997). *Machine learning* (Vol. 1). McGraw Hill.

Molano-Mazon, M., Barbosa, J., Pastor-Ciurana, J., Fradera, M., Zhang, R.-Y., Forest, J., … Yang, G. R. (2022). Neurogym: An open resource for developing and sharing neuroscience tasks. https://doi.org/10.31234/osf.io/aqc9n.

Montello, D. R. (2005). *Navigation*. Cambridge University Press.

Moskovitz, T., Miller, K., Sahani, M., & Botvinick, M. M. (2022). A unified theory of dual-process control. *arXiv preprint arXiv:2211.07036*.

Müller, S., Hollmann, N., Arango, S. P., Grabocka, J., & Hutter, F. (2021). Transformers can do Bayesian inference. *arXiv preprint arXiv:2112.10510*.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Newell, A. (1992). Unified theories of cognition and the role of soar. In *Soar: A cognitive architecture in perspective* (pp. 25–79). Springer.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Prentice Hall.

Nichol, A., Achiam, J., & Schulman, J. (2018). On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*.

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge University Press.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Ortega, P. A., Braun, D. A., Dyer, J., Kim, K.-E., & Tishby, N. (2015). Information-theoretic bounded rationality. *arXiv preprint arXiv:1512.06789*.

Ortega, P. A., Wang, J. X., Rowland, M., Genewein, T., Kurth-Nelson, Z., Pascanu, R., … Legg, S. (2019). Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 8024–8035). Curran. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pearl, J. (2009). *Causality*. Cambridge University Press.

Piloto, L. S., Weinstein, A., Battaglia, P., & Botvinick, M. (2022). Intuitive physics learning in a deep-learning model inspired by developmental psychology. *Nature Human Behaviour*, *6*(9), 1257–1267.

Pritzel, A., Uria, B., Srinivasan, S., Badia, A. P., Vinyals, O., Hassabis, D., … Blundell, C. (2017). Neural episodic control. In *International conference on machine learning* (pp. 2827–2836).

Rabinowitz, N. C. (2019). Meta-learners' learning dynamics are unlike learners'. *arXiv preprint arXiv:1905.01320*.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-Maron, G., … de Freitas, N. (2022). A generalist agent. *arXiv preprint arXiv:2205.06175*.

Rescorla, M. (2020). An improved Dutch book theorem for conditionalization. *Erkenntnis*, *87*, 1–29.

Rescorla, R. A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rich, A. S., & Gureckis, T. M. (2019). Lessons for artificial intelligence from the study of natural stupidity. *Nature Machine Intelligence*, *1*(4), 174–180.

Ritter, S., Barrett, D. G., Santoro, A., & Botvinick, M. M. (2017). Cognitive psychology for deep neural networks: A shape bias case study. In *International conference on machine learning* (pp. 2940–2949).

Ritter, S., Wang, J., Kurth-Nelson, Z., Jayakumar, S., Blundell, C., Pascanu, R., & Botvinick, M. (2018). Been there, done that: Meta-learning with episodic recall. In *International conference on machine learning* (pp. 4354–4363).

Riveland, R., & Pouget, A. (2022). Generalization in sensorimotor networks configured with natural language instructions. *bioRxiv*, 2022-02.

Rosenkrantz, R. D. (1992). The justification of induction. *Philosophy of Science*, *59*(4), 527–539.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1988). *Parallel distributed processing* (Vol. 1). IEEE.

Sanborn, A., & Griffiths, T. (2007). Markov chain Monte Carlo with people. *Advances in Neural Information Processing Systems*, *20*, 1265–1272.

Sanborn, A. N. (2017). Types of approximation for probabilistic cognition: Sampling and variational. *Brain and Cognition*, *112*, 98–101.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144.

Sanborn, A. N., & Silva, R. (2013). Constraining bridges between levels of analysis: A computational justification for locally Bayesian learning. *Journal of Mathematical Psychology*, *57*(3–4), 94–106.

Sancaktar, C., Blaes, S., & Martius, G. (2022). Curious exploration via structured world models yields zero-shot object manipulation. In A. H. Oh, A. Agarwal, D. Belgrave, & K. Cho (Eds.) , *Advances in neural information processing systems* (pp. 24170–24183). Curran Associates, Inc. https://openreview.net/forum?id=NnuYZ1el24C.

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., & Lillicrap, T. (2016). Meta-learning with memory-augmented neural networks. In *International conference on machine learning* (pp. 1842–1850).

Savage, L. J. (1972). *The foundations of statistics*. Courier.

Schaul, T., & Schmidhuber, J. (2010). Metalearning. *Scholarpedia*, *5*(6), 4650 (revision #91489). doi:10.4249/scholarpedia.4650

Schlag, I., Irie, K., & Schmidhuber, J. (2021). Linear transformers are secretly fast weight programmers. In *International conference on machine learning* (pp. 9355–9366).

Schmidhuber, J. (1987). Evolutionary principles in self-referential learning, or on learning how to learn: The meta-meta-… hook. Unpublished doctoral dissertation, Technische Universität München.

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., & Kersting, K. (2022). Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence*, *4*(3), 258–268.

Schulz, E., Bhui, R., Love, B. C., Brier, B., Todd, M. T., & Gershman, S. J. (2019). Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(28), 13903–13908.

Schulz, E., & Dayan, P. (2020). Computational psychiatry for computers. *iScience*, *23*(12), 101772.

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, *55*, 7–14.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., & Gershman, S. J. (2017). Compositional inductive biases in function learning. *Cognitive Psychology*, *99*, 44–79. doi:10.1016/j.cogpsych.2017.11.002

Schulze Buschoff, L. M., Schulz, E., & Binz, M. (2023). The acquisition of physical knowledge in generative neural networks. In *Proceedings of the 40th international conference on machine learning* (pp. 30321–30341).

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, *30*, 4080–4090.

Solomonoff, R. J. (1964). A formal theory of inductive inference. Part I. *Information and Control*, *7*(1), 1–22.

Stopper, C. M., Maric, T., Montes, D. R., Wiedman, C. R., & Floresco, S. B. (2014). Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron*, *84*(1), 177–189.

Strouse, D., McKee, K., Botvinick, M., Hughes, E., & Everett, R. (2021). Collaborating with humans without human data. *Advances in Neural Information Processing Systems*, *34*, 14502–14515.

Tamar, A., Wu, Y., Thomas, G., Levine, S., & Abbeel, P. (2016). Value iteration networks. *Advances in Neural Information Processing Systems*, *29*, 2154–2162.

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, *124*(4), 410.

Team, A. A., Bauer, J., Baumli, K., Baveja, S., Behbahani, F., Bhoopchand, A., … Zhang, L. (2023). Human-timescale adaptation in an open-ended task space. *arXiv preprint arXiv:2301.07608*.

Team, O. E. L., Stooke, A., Mahajan, A., Barros, C., Deck, C., Bauer, J., … Czarnecki, W. M. (2021). Open-ended learning leads to generally capable agents. *arXiv preprint arXiv:2107.12808*.

Tenenbaum, J. (2021). Joshua Tenenbaum's homepage. Retrieved from http://web.mit.edu/cocosci/josh.html

Thrun, S., & Pratt, L. (1998). Learning to learn: Introduction and overview. In S. Thrun & L. Pratt (Eds.), *Learning to learn* (pp. 3–17). Springer.

Todorov, E., Erez, T., & Tassa, Y. (2012). Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems* (pp. 5026–5033).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 6000–6010.

Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning. *Advances in Neural Information Processing Systems*, *29*, 3637–3645.

Wang, J. X. (2021). Meta-learning in natural and artificial intelligence. *Current Opinion in Behavioral Sciences*, *38*, 90–95.

Wang, J. X., King, M., Porcel, N., Kurth-Nelson, Z., Zhu, T., Deck, C., … Botvinick, M. (2021). Alchemy: A structured task distribution for meta-reinforcement learning. CoRR, abs/2102.02926. Retrieved from https://arxiv.org/abs/2102.02926

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., … Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, *21*(6), 860–868.

Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., … Botvinick, M. (2016). Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*.

Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A., & Cohen, J. D. (2014). Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*, *143*(6), 2074.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*(7), 1341–1390.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*(12), 915–924.

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, *22*(2), 297–306.

Yang, Y., & Piantadosi, S. T. (2022). One model for the learning of language. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(5), e2021865119.

Yu, T., Quillen, D., He, Z., Julian, R., Hausman, K., Finn, C., & Levine, S. (2020). Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on robot learning* (pp. 1094–1100).

Zednik, C., & Jäkel, F. (2016). Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, *193*(12), 3951–3985.

Zenil, H., Marshall, J. A., & Tegnér, J. (2015). Approximations of algorithmic and structural complexity validate cognitive-behavioural experimental results. *arXiv preprint arXiv:1509.06338*.

Open Peer Commentary

Challenges of meta-learning and rational analysis in large worlds

Margherita Calderan[a] and Antonino Visalli[b]*

[a]Department of Developmental Psychology and Socialisation, University of Padova, Italy and [b]IRCCS San Camillo Hospital, Venice, Italy
margherita.calderan@phd.unipd.it
antonino.visalli@hsancamillo.it

*Corresponding author.

doi:10.1017/S0140525X24000128, e148

Abstract

We challenge Binz et al.'s claim of meta-learned model superiority over Bayesian inference for large world problems. While comparing Bayesian priors to model-training decisions, we question meta-learning feature exclusivity. We assert no special justification for rational Bayesian solutions to large world problems, advocating exploring diverse theoretical frameworks beyond rational analysis of cognition for research advancement.

Binz et al. (argument 2) advocate for the superiority of meta-learned models over Bayesian inference for addressing large world problems (Savage, 1972). Our commentary aims to question some perceived fallacies in their arguments.

First, although we recognize that "*Identifying the correct set of assumptions becomes especially challenging once we deal with more complex problems*," we point out that meta-learned models also require specific assumptions. Examples are the selection of samples from the data-generating distribution, choice of the optimizer, weight initializations, or constraints to mimic bounded rationality. These decisions, too, can be conceived as priors and require a certain level of justification. Binz et al. explicitly emphasized the importance of appropriately making these choices (sect. 4, "Intricate Training Processes"). Bayesian or not, prior knowledge is a necessary condition for both modeling procedures. As a consequence, we contend that both Bayesian and meta-learned models present similar challenges from a rational perspective. Therefore, why should it be "*hard to justify*" prior assumptions for Bayesian models and not for meta-learned models? For instance, one could reconsider the critiques moved to Lucas, Griffiths, Williams, and Kalish (2015). To account for the bias toward expecting linear relationships between continuous variables, the authors assigned lower prior probabilities to quadratic and radial relationships as compared to linear ones (Lucas et al., 2015). Binz et al. pose the issue that the chosen prior might not reflect all the functions (and the associated probability). However, similar concerns arise in the context of meta-learned models. What justifications exist for the selection of training data? How does one determine which functions to employ in the tasks used for training the model? Even more, on which tasks the model should be trained? Are these decisions easier to justify from a rational perspective as compared to the Bayesian counterpart? If the definition of the priors is considered a main obstacle of Bayesian inference to

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

large world problems, a similar challenge extends to the decisions mentioned above, which determine the initial parameterization of meta-learned models and could be conceived as equivalent to a "prior" (Griffiths et al., 2019). Finally, if it is the impossibility to "*have access to a prior or a likelihood*" the main obstacle to large world problems, what is "*the unique feature of meta-learned models*" compared to other Bayesian methods that can construct their own empirical priors (e.g., hierarchical models and empirical Bayes; Friston & Stephan, 2007) or that bypass the evaluation of the likelihood function (e.g., approximate Bayesian computation: Beaumont, 2010; likelihood-free inference: Papamakarios, Nalisnick, Rezende, Mohamed, & Lakshminarayanan, 2021; simulation-based inference: Cranmer, Brehmer, & Louppe, 2020)?

Second, it should be noted that the "meta-learning" feature is not exclusive of meta-learned models in machine learning, but it can be achieved using hierarchical Bayesian models (Grant, Finn, Levine, Darrell, & Griffiths, 2018; Griffiths et al., 2019; Kemp, Perfors, & Tenenbaum, 2007; Li, Callaway, Thompson, Adams, & Griffiths, 2023). Hence, if meta-learning is taken as an argument to enable computational models to face large world problems, it cannot be used as an argument in favor of meta-learned models over hierarchical Bayesian inference.

Putting together our first two concerns, we think that a more fair comparison between meta-learned models (as defined in the target article) and hierarchical (approximate) Bayesian models would have been necessary to assert that meta-learned models contain "*unique*" features to address large world problems.

A further concern regards meta-learning as a solution for large world problems. Following Binmore (2007), the distinction between small and large worlds can be interpreted as making decisions under risk or uncertainty, respectively. In the first case, decision makers know all contingencies of the problem and fully apply the Bayes' rule to make the optimal decision. Large world problems are situations characterized by uncertainty about the causes and the likelihood of the events. In other terms, large world problems can be conceived as situations in which environmental assumptions previously acquired do not hold. However, if meta-learned models need to be retrained when environmental assumptions differ from the training, it follows that the use of meta-learned models can be justified only in small worlds, where previous knowledge can be used to make choices.

Finally, it should be highlighted that the target article grounds meta-learned models on the rational analysis framework (Anderson, 1991) given their property of approximate Bayes optimal solutions. However, Savage's and Binmore's argument was that there is no special justification for rational Bayesian solutions to large world problems. In our opinion, if one wants to hold with this rational perspective, neither Bayesian nor meta-learned models can be considered idoneous to model decision making under uncertainty. However, a possible way out of this impasse can come from psychological and cognitive research fields that have investigated decision making under uncertainty. Theoretical frameworks like the free-energy principle (Friston et al., 2023) or reinforcement learning (Dimitrakakis & Ortner, 2022; Kochenderfer, 2015) have investigated how learning under uncertainty occurs and it is used to construct beliefs that guide decisions in situations where causes of the event are unknown. In our opinion, implementing ideas from these frameworks in the models can be a promising way to solve large world problems.

In conclusion, we do not think that Binz et al. have provided convincing support for the claim "*The ability to construct Bayes-optimal learning algorithms for large world problems is a unique feature of the meta-learning framework.*" We suggest that grounding on the rational analysis of cognition framework is not sufficient for modeling decisions in large worlds, and that exploring and integrating other theoretical frameworks could offer valuable insights to advance their research program.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interest. None.

References

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*(3), 471–485.

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 379–406.

Binmore, K. (2007). Rational decisions in large worlds. *Annales d'Economie et de Statistique*, *86*, 25–41.

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062.

Dimitrakakis, C., & Ortner, R. (2022). *Decision making under uncertainty and reinforcement learning: Theory and algorithms* (Vol. 223). Springer Nature.

Friston, K. J., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G. A., & Parr, T. (2023). The free energy principle made simpler but not too simple. *Physics Reports*, *1024*, 1–29.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*, 417–458.

Grant, E., Finn, C., Levine, S., Darrell, T., & Griffiths, T. (2018). Recasting gradient-based meta-learning as hierarchical Bayes. arXiv preprint arXiv:1801.08930.

Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., & Lieder, F. (2019). Doing more with less: Meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, *29*, 24–30.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Kochenderfer, M. J. (2015). *Decision making under uncertainty: Theory and application*. MIT press.

Li, M. Y., Callaway, F., Thompson, W. D., Adams, R. P., & Griffiths, T. L. (2023). Learning to learn functions. *Cognitive Science*, *47*(4), e13262.

Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review*, *22*(5), 1193–1215.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research*, *22*(1), 2617–2680.

Savage, L. J. (1972). *The foundations of statistics*. Courier Corporation.

Meta-learning modeling and the role of affective-homeostatic states in human cognition

Ignacio Cea[a,b,c*]

[a]Faculty of Religious Sciences and Philosophy, Temuco Catholic University, Temuco, Chile; [b]Center for Research, Innovation and Creation, Temuco Catholic University, Temuco, Chile and [c]Universidad Alberto Hurtado, Faculty of Philosophy and Humanities, Philosophy Department, Santiago, Chile
icea@uct.cl
https://igneocj.wixsite.com/ignacio-cea

*Corresponding author.

doi:10.1017/S0140525X24000098, e149

Abstract

The meta-learning framework proposed by Binz et al. would gain significantly from the inclusion of affective and homeostatic elements, currently neglected in their work. These components are crucial as cognition as we know it is profoundly influenced by affective states, which arise as intricate forms of homeostatic regulation in living bodies.

Binz et al. offer a very promising research program based on meta-learning to advance our understanding of cognition. Nonetheless, their proposal could greatly benefit from integrating affective and homeostatic elements, which have been traditionally underrepresented in cognitive science and are totally absent in their article as well. This integration is predicated on the premise that cognitive processes in general are profoundly influenced by affective states (e.g., emotions and moods), which arise as intricate forms of homeostatic regulation in living organisms.

First, there is ample evidence from psychology and affective neuroscience showing the pervasive influence of affective states on cognitive processes (Cea, 2023; Clore & Schiller, 2018). When someone feels affectively moved by new information, she will more likely attend to it and think about it for an extended duration, compared to neutral information (Manns & Bass, 2016). People in positive moods tend to engage in more creative thinking and learn subjects meaningfully (Ormrod, Anderman, & Anderman, 2019), while those feeling sadness or frustration are prone to a shallower learning (Ahmed, Van der Werf, Kuyper, & Minnaert, 2013). This is related to negative emotions being accompanied by cortisol release and the fight-or-flight response that can suppress the prefrontal cortex, thereby hindering higher cognition (Brackett, 2019). Hence, it is reasonable that emotion regulation abilities are the strongest predictors of academic achievement in high-school students (Di Fabio & Palazzeschi, 2009).

Also, our affective states shape what and how we perceive (Barrett, 2017; Cea & Martínez-Pernía, 2023). According to the affective information principle, our feelings signal the value and urgency of any perceived object (Clore & Schiller, 2018). Moreover, by altering people's affects, researchers can sway perceptions, for example, making a beverage seem appealing or distasteful (Berridge & Winkielman, 2003), and people friendly or mean (Li, Moallem, Paller, & Gottfried, 2007). A key brain area involved is the orbitofrontal cortex, which integrates sensory and affective information, ensuring that our perceptions are always imbued with affect (Barrett & Bar, 2009).

Concerning attention, positive emotions generally broaden it, leading to a global focus (Fredrickson & Branigan, 2005), whereas negative emotions narrow it, fostering detail-oriented processing (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & Van Ijzendoorn, 2007). Also, emotionally charged stimuli are shown to capture attention more effectively than neutral ones, both in terms of speed and focus (Hajcak, Jackson, Ferri, & Weinberg, 2018). This can cause perceptual interference, making subsequent neutral stimuli less noticeable (Wang, Kennedy, & Most, 2012), a process associated with emotional allocation of attentional resources as indicated by the late positive potential in the parietal lobe (Hajcak et al., 2018).

Concerning memory, William James claimed that "an experience may be so exciting emotionally as almost to leave a scar upon the cerebral tissues" (James, 1890, p. 670). People tend to remember information with emotional significance more easily than neutral material (Phelps & Sharot, 2008). Memories associated with strong emotional arousal are recalled more vividly (Schaefer & Philippot, 2005) and with greater detail in certain respects compared to neutral memories (Kensinger & Schacter, 2018). Neurobiologically, the reciprocal interactions between the amygdala and the hippocampus are considered essential for this (McGaugh, 2013).

Concerning the homeostatic roots of affect, Seth and Barrett have independently suggested that affect arises from the brain's inferences about the causes of internal body signals to regulate physiological states (e.g., sugar levels, heart-beat, etc.), ensuring survival based on past experiences (Barrett, 2017; Barrett & Simmons, 2015; Seth, 2021; Seth & Tsakiris, 2018). Hence, according to them, our feelings would then be expressions of our current and future degree of success or failure in staying alive. In this way, moods and emotions would be intimately linked to our bodily nature and homeostatic needs.

This core idea of feelings being rooted in homeostatic regulation in vulnerable systems has been applied to robotics and artificial intelligence. Man and Damasio (2019) propose a novel class of soft robots that incorporate physical vulnerability and self-regulation akin to living organisms. They hypothesize that this would allow them to develop motivations and evaluations reminiscent of feelings in humans, potentially leading to more intelligent interactions with their environments. Similarly, Bronfman, Ginsburg, and Jablonka (2021) propose that feelings may arise in artificial systems constructed with soft materials through the development of homeostatic, self-preservation mechanisms that could allow them to instantiate an open-ended domain-general form of learning, what they call unlimited associative learning. Finally, Yoshida (2017) introduces a reinforcement learning model where agents learn to survive by regulating critical variables like energy levels, using a reward system rooted in homeostatic principles, leading to adaptive behavior. Importantly, all proposals emphasize that the possibility of engineering sentient artificial systems depends on having vulnerable bodies that, akin to ourselves, need to be constantly sensed and regulated to remain integral, and that this would enhance the machines' cognitive capacities.

To conclude, I would like to suggest some potential benefits of incorporating affective and homeostatic elements into the meta-learning research program: (i) *Enhanced adaptability*: By incorporating these elements into the meta-learning algorithms, the resulting computational models may better simulate the adaptability of human cognition, like the ability to adjust learning strategies to changing environmental and internal states; (ii) *richer contextual understanding*: Incorporating affective-homeostatic elements in learning-to-learn processes can result in a deeper understanding of how emotionally salient contexts influence cognition; (iii) *improved learning efficiency*: Affective-homeostatic signals can guide attention and memory processes, leading to more efficient learning. Meta-learning algorithms that incorporate affective-homeostatic signals or mechanisms may achieve higher efficiency in adapting learning algorithms to new tasks or to new information; (iv) *more realistic simulations*: By incorporating affect and homeostasis, meta-learned models may more accurately simulate human cognition, which is inherently influenced by affective-bodily states; and (v) *cross-domain generalization*: The integration of affective-homeostatic states may facilitate better generalization across different cognitive domains, as affect and bodily regulation often play a role in a wide range of cognitive tasks, from decision making to social interactions.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

In sum, I encourage Binz et al. to consider the beneficial prospects of incorporating these elements into their proposed research program, so as to acknowledge the intertwined nature of affect, bodily homeostasis, and human cognition.

Acknowledgements. I am very grateful to Kingson Man for his constructive comments on an earlier version of this commentary. I am also very grateful to Thomas Wachter for his encouraging appreciation of my previous work on affect, homeostasis, and cognition.

Financial support. This research was funded by ANID-Fondecyt Postdoctoral grant #3210707 and the Center for Research, Innovation and Creation, Temuco Catholic University.

Competing interest. None.

References

Ahmed, W., Van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, self-regulated learning, and achievement in mathematics: A growth curve analysis. *Journal of Educational Psychology*, *105*(1), 150.

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F., & Bar, M. (2009). See it with feeling: Affective predictions during object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364* (1521), 1325–1334.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429.

Berridge, K., & Winkielman, P. (2003). What is an unconscious emotion?(The case for unconscious "liking"). *Cognition and Emotion*, *17*(2), 181–211. https://doi.org/10.1080/02699930302289

Brackett, M. (2019). *Permission to feel: Unlocking the power of emotions to help our kids, ourselves, and our society thrive*. Celadon Books.

Bronfman, Z., Ginsburg, S., & Jablonka, E. (2021). When will robots be sentient? *Journal of Artificial Intelligence and Consciousness*, *8*(02), 183–203.

Cea, I. (2023). The somatic roots of affect: Toward a body-centered education. In P. Fossa, & C. Cortés-Rivera (Eds.), *Affectivity and learning: Bridging the gap between neurosciences, cultural and cognitive psychology* (pp. 555–583). Springer.

Cea, I., & Martínez-Pernía, D. (2023). Continuous organismic sentience as the integration of core affect and vitality. *Journal of Consciousness Studies*, *30*(3–4), 7–33. https://doi.org/https://doi.org/10.53765/20512201.30.3.007

Clore, G., & Schiller, A. (2018). New light on the affect-cognition connection. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions*, (4th ed., pp. 532–546). The Guilford Press.

Di Fabio, A., & Palazzeschi, L. (2009). An in-depth look at scholastic success: Fluid intelligence, personality traits or emotional intelligence? *Personality and Individual Differences*, *46*(5–6), 581–585.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion*, *19*(3), 313–332.

Hajcak, G., Jackson, F., Ferri, J., & Weinberg, A. (2018). Emotion and attention. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions*, (4th ed., pp. 595–609). The Guilford Press.

James, W. (1890). *The principles of psychology* (Vol. 1, Issue n/a). Dover Publications.

Kensinger, E. A., & Schacter, D. L. (2018). Memory and emotion. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions*, (4th ed., pp. 564–578). The Guilford Press.

Li, W., Moallem, I., Paller, K. A., & Gottfried, J. A. (2007). Subliminal smells can guide social preferences. *Psychological Science*, *18*(12), 1044–1049.

Man, K., & Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence*, *1*(10), 446–452.

Manns, J. R., & Bass, D. I. (2016). The amygdala and prioritization of declarative memories. *Current Directions in Psychological Science*, *25*(4), 261–265.

McGaugh, J. L. (2013). Making lasting memories: Remembering the significant. *Proceedings of the National Academy of Sciences*, *110*(supplement_2), 10402–10407.

Ormrod, J. E., Anderman, E. M., & Anderman, L. H. (2019). *Educational psychology: Developing learners*, (10th ed.). Pearson.

Phelps, E. A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science*, *17*(2), 147–152.

Schaefer, A., & Philippot, P. (2005). Selective effects of emotion on the phenomenal characteristics of autobiographical memories. *Memory*, *13*(2), 148–160.

Seth, A. (2021). *Being you: A new science of consciousness*. Penguin.

Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, *22*(11), 969–981. https://doi.org/10.1016/j.tics.2018.08.008

Wang, L., Kennedy, B. L., & Most, S. B. (2012). When emotion blinds: A spatiotemporal competition account of emotion-induced blindness. *Frontiers in Psychology*, *3*, 438.

Yoshida, N. (2017). Homeostatic agent for general environment. *Journal of Artificial General Intelligence*, *8*(1), 1.

Quantum Markov blankets for meta-learned classical inferential paradoxes with suboptimal free energy

Kevin B. Clark[a–k,*]

[a]Cures Within Reach, Chicago, IL, USA; [b]Felidae Conservation Fund, Mill Valley, CA, USA; [c]Campus and Domain Champions Program, Multi-Tier Assistance, Training, and Computational Help (MATCH) Track, National Science Foundation's Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS); [d]Expert Network, Penn Center for Innovation, University of Pennsylvania, Philadelphia, PA, USA; [e]Network for Life Detection (NfoLD), NASA Astrobiology Program, NASA Ames Research Center, Mountain View, CA, USA; [f]Multi-Omics and Systems Biology & Artificial Intelligence and Machine Learning Analysis Working Groups, NASA GeneLab, NASA Ames Research Center, Mountain View, CA, USA; [g]Frontier Development Lab, NASA Ames Research Center, Mountain View, CA, USA; [h]SETI Institute, Mountain View, CA, USA; [i]Peace Innovation Institute, The Hague 2511, Netherlands & Stanford University, Palo Alto, CA, USA; [j]Shared Interest Group for Natural and Artificial Intelligence (sigNAI), Max Planck Alumni Association, Berlin, Germany and [k]Biometrics and Nanotechnology Councils, Institute for Electrical and Electronics Engineers (IEEE), New York, NY, USA
kbclarkphd@yahoo.com
www.linkedin.com/pub/kevin-clark/58/67/19a
https://access-ci.org/

*Corresponding author.

doi:10.1017/S0140525X24000244, e150

Abstract

Quantum active Bayesian inference and quantum Markov blankets enable robust modeling and simulation of difficult-to-render natural agent-based classical inferential paradoxes interfaced with task-specific environments. Within a non-realist cognitive completeness regime, quantum Markov blankets ensure meta-learned irrational decision making is fitted to explainable manifolds at optimal free energy, where acceptable incompatible observations or temporal Bell-inequality violations represent important verifiable real-world outcomes.

Applying a rational analysis framework of cognition, Binz et al. resolutely embrace the escalating use of meta-learning to re-construct Bayes optimal-learning algorithms and solutions to explain the metaphysical relationship of mind to environment. Such logicomathematical descriptions of cognition non-trivially approximate properties of mind, including agent-based decision processes, to that of ideal statistical inference and the structure of natural tasks and environments. The authors nonetheless fail to satisfactorily introduce Markov blankets, which would expand their cognitive construct beyond standard applications of analytical and numerical tools to demark relations represented in

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

directed Bayesian networks or graphs and variational probabilistic inference. Markov or Pearl blankets, coined by Bruineberg, Dolega, Dewhurst, and Baltieri (2021) to acknowledge the original decades-old epistemic concept developed by Pearl (1988), earned deserving attention for their utility in decision sciences, offering tractable optimization methods to identify, partition, and understand groups of marginally and conditionally (in)dependent variables associated with complex systems and causal predictions and attributions. In more recent years, however, a trend started by Friston and coworkers (e.g., Hipólito et al., 2021; Friston et al., 2021; Ramstead, Badcock, & Friston, 2018) promotes extension of Markov blankets within a free-energy, active-inference framework to describe the physical interface and reciprocal interactions between agents and their environments. These so-called Friston blankets, a term also coined by Bruineberg et al. (2021) to differentiate their usage from Pearl blankets, cleverly articulate embedded philosophical axioms of mind, body, and environment without fully capturing the logicomathematical rigor and cogency found in typical uses of Pearl blankets and Bayesian inference. Taking a reasonably conventional position on the state-of-art of Friston blankets, theorists supportive of free-energy models must either accept Markov blankets as formal technical innovations that enable practical worthwhile science in absence of known compelling philosophical conclusions about the nature of cognition and life or as dubious mathematics-driven metaphysics interpretations of reality with promising high-impact implications for elucidating cognition and life upon separate experimental biophysical validation. The merits of such a statement seemly convey a fair, albeit critical, edict to the scientific community – one that perhaps discouraged Binz et al. from clarifying Markov blankets and leaves the impression that good productive science enlisting Markov blankets as instruments for inference returns only theoretically mundane findings about cognition and, possibly, life. But, that is not the case and the authors' well-structured arguments may fool readers into believing that this position is true since Binz et al. disappointingly content themselves with only discussing classical Bayesian inference and non-paradoxical cognition, maybe because Friston and coworkers also never stray beyond a classical formulation of their free-energy principle for active inference.

Weaknesses in Binz et al.'s narrow perspective on meta-leaning and cognition may be contrasted and reshaped by exciting findings produced with quantum decision theory, a strict valid quantum-statistical approach capable of defining probabilistic human inference unconfined by the physical mechanical world (Aerts & Aerts, 1995; Aerts, Broakaert, Gabara, & Sozzo, 2016; Ashtiani & Azgomi, 2015; Busemeyer & Bruza, 2011; Busemeyer, Wang, & Lambert-Mogiliansky, 2009; Clark, 2011, 2012, 2014b, 2015, 2017; Pinto Moreira, Fell, Dehdashti, Bruza, & Wichert, 2020; Pothos & Busemeyer, 2013). Quantum decision theory, with the aid of quantum networks or graphs, quantum Bayesian inference, and other computational features, demonstrates robust successes in modeling and simulating difficult-to-render cognitive phenomena overlooked by Binz et al., especially causal judgment errors or paradoxes unaccounted for by classical decision theory, including conjunctive and disjunctive fallacies, the Allais paradox, and the Ellsberg or planning paradox (Atmanspacher & Römer, 2012; Busemeyer, Pothos, Franco, & Trueblood, 2011; Clark, 2021b; Favre, Wittwer, Heinimann, Yukalov, & Sornette, 2016; Moreira & Wichert, 2016a, 2016b, 2018b; Pothos & Busemeyer, 2009, 2013). The great explanatory power of quantum decision networks (Bianconi, 2002a, 2002b, 2003; Bianconi & Barabási, 2001; Li, Iqbal, Perc, Chen, & Abbott, 2013) permits expression of quantum Markov chains and blankets (Brandao, Piani, & Horodecki, 2015; Moreira & Wichert, 2018a; Qi & Ranard, 2021; Sutter, 2018; Wichert, Pinto Moreira, & Bruza, 2020), which bound suboptimal free-energy classical inferential paradoxes from optimal free-energy quantum inferential solutions. In this context of meta-learned cognitive inference, both quantum Bayesian inference and Markov blankets provide epistemic tools, analogous to variational Bayesian inference and Pearl blankets, to legitimately approximate irrational human decision making and cognition within a cognitive-completeness constraint. That constraint, not considered by Binz et al., strongly limits degrees of freedom for emergence of meta-learned subjective physical reality (cf. Blume-Kohout & Zurek, 2006; Brandao et al., 2015; Clark, 2014a, 2017, 2019, 2020, 2021b, 2023; Yearsley & Pothos, 2014), a scenario which helps affirm the idea that Markov blankets may yield meaningful philosophical conclusions regarding the nature of cognition and life.

Cognitive completeness (Tressoldi, Maier, Buechner, & Khrennikov, 2015; Yearsley & Pothos, 2014) encapsulates a black-box approach that isolates any studied cognitive system from the formidable environment-significant measurement problem of quantum mechanics. Some scientists insist the scalable neurophysiological contents of this black box map onto classical or quantum cognitive states relevant to particular sets of respective rational or irrational decisions and their corresponding outcome probabilities (Clark, 2017, 2021a; Wang & Busemeyer, 2015; Yearsley & Pothos, 2014). For example, if one abandons the constraint of cognitive realism – the assertion that all (meta-learned) cognitive events emerge from classical deterministic neurophysiology – then cognitive systems and their decisional outcomes may be completely described by dimensions or sets of similarity classes on the set of all probability distributions over deterministic and indeterminate (or stochastic) neuropsychological variables and associated environmental settings. Further, arguably more fundamental cognitive completeness and Markov blanket partitions imply non-disturbing measurements should be non-invasive on brain physiology and cognition, with disturbing measurements affecting outcomes of hidden neurophysiological and cognitive variables in a quantum-sensitive manner. Disturbing measurements violate temporal Bell or Leggett–Garg inequalities, signifying violations of classical (meta-learned) cognition, such as Markov-blanketed inferential paradoxes at suboptimal free energy. Restructuring these cognitive phenomena within quantum decision theory and quantum Markov blankets creates opportunities for irrational decision making to be organized into explainable manifolds at optimal free energy, tantamount to quantum active Bayesian inference where observable incompatibility or inequality violations are acceptable (Clark, 2021a). Such theoretical elegance in describing complex cognition forces classical and quantum aspects of brain structure and function into a newer realm of logicomathematical formalism with verifiable, important real-world consequences.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interest. None.

References

Aerts, D., & Aerts, S. (1995). Applications of quantum statistics in psychological studies of decision processes. *Foundations of Science*, *1*, 85–97.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Aerts, D., Broakaert, J., Gabara, L., & Sozzo, S. (Eds.) (2016). *Quantum structures in cognitive and social sciences*. Frontiers Research Topics Ebook. Frontiers Media SA. ISBN 978-2-88919-876-4.
Ashtiani, M., & Azgomi, M. A. (2015). A survey of quantum-like approaches to decision making and cognition. *Mathematical Social Sciences*, *75*, 49–80.
Atmanspacher, H., & Römer, H. (2012). Order effects in sequential measurements of noncommuting psychological observables. *Journal of Mathematical Psychology*, *56*, 274–280.
Bianconi, G. (2002a). Growing Cayley trees described by a Fermi distribution. *Physical Review E: Statistical, Nonlinear, Soft Matter Physics*, *66*, 036116.
Bianconi, G. (2002b). Quantum statistics in complex networks. *Physical Review E: Statistical, Nonlinear, Soft Matter Physics*, *66*, 056123.
Bianconi, G. (2003). Size of quantum networks. *Physics Review E*, *67*(5, Pt 2), 056119.
Bianconi, G., & Barabási, A.-L. (2001). Bose-Einstein condensation in complex networks. *Physics Review Letters*, *86*, 5632–5635.
Blume-Kohout, R., & Zurek, W. H. (2006). Quantum Darwinism: Entanglement, branches, and the emergent classicality of redundantly stored quantum information. *Physical Review A*, *73*(6), 062310.
Brandao, F. G. S. L., Piani, M., & Horodecki, P. (2015). Generic emergence of classical features in quantum Darwinism. *Nature Communications*, *6*, 7908.
Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2021). The emperor's new Markov blankets. *Behavior and Brain Sciences*, *45*, e183.
Busemeyer, J. R., & Bruza, P. (2011). *Quantum models of cognition and decision making*. Cambridge University Press. ISBN 13 978-1107419889.
Busemeyer, J. R., Pothos, E., Franco, R., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment "errors". *Psychological Reviews*, *118*, 193–218.
Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). Empirical comparison of Markov and quantum models of decision making. *Journal of Mathematical Psychology*, *53*, 423–433.
Clark, K. B. (2011). Live soft-matter quantum computing. In E. C. Salander (Ed.), *Computer search algorithms* (pp. 1–24). Nova Science Publishers, Inc. ISBN 978-1-61122-527-3.
Clark, K. B. (2012). A statistical mechanics definition of insight. In A. G. Floares (Ed.), *Computational intelligence* (pp. 139–162). Nova Science Publishers, Inc. ISBN 978-1-62081-901-2.
Clark, K. B. (2014a). Basis for a neuronal version of Grover's quantum algorithm. *Frontiers in Molecular Neuroscience*, *7*, 29.
Clark, K. B. (2014b). Evolution of affective and linguistic disambiguation under social eavesdropping pressures. *Behavioral and Brain Sciences*, *37*(6), 551–552.
Clark, K. B. (2015). Insight and analysis problem solving in microbes to machines. *Progress in Biophysics and Molecular Biology*, *119*, 183–193.
Clark, K. B. (2017). Cognitive completeness of quantum teleportation and superdense coding in neuronal response regulation and plasticity. *Proceedings of the Royal Society B: Biological Sciences*. eLetter. https://royalsocietypublishing.org/doi/suppl/10.1098/rspb.2013.3056
Clark, K. B. (2019). Unpredictable homeodynamic and ambient constraints on irrational decision making of aneural and neural foragers. *Behavioral and Brain Sciences*, *42*, e40.
Clark, K. B. (2020). Digital life, a theory of minds, and mapping human and machine cultural universals. *Behavioral and Brain Sciences*, *43*, e98.
Clark, K. B. (2021a). Quantum decision corrections for the neuroeconomics of irrational movement control and goal attainment. *Behavioral and Brain Sciences*, *44*, e127.
Clark, K. B. (2021b). Classical and quantum information processing in aneural to neural cellular decision making on Earth and perhaps beyond. *Bulletin of the American Astronomical Society*, *53*(4), 33.
Clark, K. B. (2023). Neural field continuum limits and the partitioning of cognitive-emotional brain networks. *Biology*, *12*(3), 352.
Favre, M., Wittwer, A., Heinimann, H. R., Yukalov, V. I., & Sornette, D. (2016). Quantum decision theory in simple risky choices. *PLoS ONE*, *11*(12), e0168045.
Friston, K. J., Parr, T., Hipolito, I., Magrou, L., & Razi, A. (2021). Parcels and particles: Markov blankets in the brain. *Network Neuroscience*, *5*(1), 211–251.
Hipolito, I., Ramstead, M. J. D., Convertino, L., Bhat, A., Friston, K., & Parr, T. (2021). Markov blankets in the brain. *Neuroscience and Biobehavioral Reviews*, *125*, 88–97.
Li, Q., Iqbal, A., Perc, M., Chen, M., & Abbott, D. (2013). Coevolution of quantum and classical strategies on evolving random networks. *PLoS ONE*, *8*(7), e68423.
Moreira, C., & Wichert, A. (2016a). Quantum-like Bayesian networks for modeling decision making. *Frontiers in Psychology*, *7*, 11.
Moreira, C., & Wichert, A. (2016b). Quantum probabilistic models revisited: The case of disjunction effects in cognition. *Frontiers in Psychology*, *4*, 26.
Moreira, C., & Wichert, A. (2018a). Are quantum-like Bayesian networks more powerful than classical Bayesian networks? *Journal of Mathematical Psychology*, *28*, 73–83.
Moreira, C., & Wichert, A. (2018b). Introducing quantum-like influence diagrams for violations of the Sure Thing Principle. *International symposium on quantum interactions* Q1 2018: 91–108.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publishers Inc. ISBN 1558604790.
Pinto Moreira, C., Fell, L., Dehdashti, S., Bruza, P., & Wichert, A. (2020). Towards a quantum-like cognitive architecture for decision-making. *Behavioral and Brain Sciences*, *43*, e171–e178.
Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of "rational" decision theory. *Proceedings of Biological Sciences*, *276*, 2171–2178.
Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, *36*, 255–327.
Qi, X.-L., & Ranard, D. (2021). Emergent classicality in general multipartite states and channels. *Quantum*. https://arxiv.org/abs/2001.01507
Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, *24*, 1–16.
Sutter, D. (2018). *Approximate quantum Markov chains*. Springer. ISBN 978-3-319-78732-9.
Tressoldi, P. E., Maier, M. A., Buechner, V. L., & Khrennikov, A. (2015). A macroscopic violation of no-signaling in time inequalities? How to test temporal entanglement with behavioral observables. *Frontiers in Psychology*, *6*, 1061.
Wang, Z., & Busemeyer, J. (2015). Reintroducing the concept of complementarity to psychology. *Frontiers in Psychology*, *6*, 1822.
Wichert, A., Pinto Moreira, C., & Bruza, P. (2020). Balanced quantum-like Bayesian networks. *Entropy*, *22*(2), 1701–1726.
Yearsley, J. M., & Pothos, E. M. (2014). Challenging the classical notion of time in cognition: A quantum perspective. *Proceedings of the Royal Society London B: Biological Sciences*, *281*, 20133056.

Meta-learning goes hand-in-hand with metacognition

Chris Fields[a] and James F. Glazebrook[b,c*]

[a]Allen Discovery Center, Tufts University, Medford, MA, USA; [b]Department of Mathematics and Computer Science, Eastern Illinois University, Charleston, IL, USA and [c]Adjunct Faculty (Mathematics), University of Illinois at Urbana-Champaign, Urbana, IL, USA
fieldsres@gmail.com
jfglazebrook@eiu.edu
https://chrisfieldsresearch.com
https://faculty.math.illinois.edu/glazebro/

*Corresponding author.

doi:10.1017/S0140525X24000256, e151

Abstract

Binz et al. propose a general framework for meta-learning 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.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

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., Cox, 2005; Flavell, 1979; Shea & Frith, 2019). 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 et al., 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; Friston et al., 2023; Ramstead et al., 2022); indeed the FEP itself is just a classical limit of the principle of unitarity, that is, of conservation of information (Fields et al., 2023; Fields, Friston, Glazebrook, & Levin, 2022). 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, Manika, Hoel, & Levin, 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. Samek, Montavon, Lapuschkin, Anders, & Müller, 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 & 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 & Chater, 2016), the relevance of heuristics (Gigerenzer & Gaissmaier, 2011; cf. Fields & Glazebrook, 2020), and how heuristics, biases, and confabulation limit reportable self-knowledge (Fields, Glazebrook, & 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.

Financial support. The authors have received no funding towards this contribution.

Competing interests. None.

References

Biswas, S., Clawson, W., & Levin, M. (2022). Learning in transcriptional network models: Computational discovery of pathway-level memory and effective interventions. *International Journal of Molecular Sciences*, *24*, 285.

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

Chater, N. (2018). *The mind is flat. The remarkable shallowness of the improvising brain*. Yale University Press.

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

Fields, C., Fabrocini, F., Friston, K. J., Glazebrook, J. F., Hazan, H., Levin, M., & Marcianò, A. (2023). Control flow in active inference systems, part I: Classical and quantum formulations of active inference. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, *9*, 235–245.

Fields, C., Friston, K. J., Glazebrook, J. F., & Levin, M. (2022). A free energy principle for generic quantum systems. *Progress in Biophysics and Molecular Biology*, *173*, 36–59.

Fields, C., & Glazebrook, J. F. (2020). Do process-1 simulations generate the epistemic feelings that drive process-2 decision making? *Cognitive Processing*, *21*, 533–553.

Fields, C., Glazebrook, J. F., & Levin, M. (2024). Principled limitations on self-representations for generic physical systems. *Entropy*, *26*(3), 194.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *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., & Parr, T. (2023) Path integrals, particular kinds, and strange things. *Physics of Life Reviews*, *47*, 35–62.

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

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Glahn, D. C., Winkler, A. M., Kochunov, P., & Blangero, J. (2010) Genetic control over the resting brain. *Proceedings of the National Academy of Sciences of the USA*, *10*(7), 1223–1228.

Kuchling, F., Fields, C., & 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. *Frontiers in Systems Neuroscience*, *16*, 768201.

Levin, M. (2023). Darwin's agential materials: Evolutionary implications of multiscale competency in developmental biology. *Cellular and Molecular Life Sciences*, *80*(6), 142.

Ramstead, M. J.,Sakthivadivel, D. A. R., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., … Friston, K. J. (2022). On Bayesian mechanics: A physics of and by beliefs. *Interface Focus*, *13*(2923), 20220029.

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

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893.
Shea, N., & Frith, C. D. (2019). The global workspace needs metacognition. *Trends in Cognitive Sciences*, *23*, 560–571.
Taylor, J. E. T., & Taylor, G. W. (2021). Artificial cognition: How experimental psychology can help generate artificial intelligence. *Psychonomic Bulletin and Review*, *28*, 454–475.
Wang, J. X. (2021). Meta-learning in artificial and natural intelligence. *Current Opinion in Behavioral Sciences*, *38*, 90–95.

The meta-learning toolkit needs stronger constraints

Erin Grant*

UCL Gatsby Unit, Sainsbury Wellcome Centre, University College London, London, UK
erin.grant@ucl.ac.uk
https://eringrant.github.io/

*Corresponding author.

doi:10.1017/S0140525X24000104, e152

Abstract

The implementation of meta-learning targeted by Binz et al. inherits benefits and drawbacks from its nature as a connectionist model. Drawing from historical debates around bottom-up and top-down approaches to modeling in cognitive science, we should continue to bridge levels of analysis by constraining meta-learning and meta-learned models with complementary evidence from across the cognitive and computational sciences.

Meta-learning as a model allows researchers to posit how human and other biological learning systems might learn from experience in a structured manner, including by relating experiences across timescales or latent causes non-uniformly. Meta-learning as a tool allows researchers to posit flexible and data-driven learning algorithms as computational models of human learning than those are readily expressed by machine learning algorithms such as gradient descent with canonical parameters, or inference in a Bayesian model in which exact inference is tractable. These senses of a "meta-learning" and a "meta-learned" model align with the dichotomy employed in Binz et al.

Meta-learning in both senses and using the implementation focused on in Binz et al. – a recurrent neural network – further inherits characteristics of connectionism: Universal approximation, ease of specification, manipulability (including of complexity), and integration of neuroscientific findings, which Binz et al. rightly note as positives. However, this implementation of meta-learning also inherits the challenges of a connectionist approach: Lack of interpretability (the ease with which humans can understand the workings and outputs of a system) and controllability (the ability to modify a model's behavior or learning process to achieve specific outcomes).

These benefits and drawbacks of the bottom-up, emergentist approach of connectionism have been discussed at length, including in this journal (Smolensky, 1988). As a result of these discussions, a common ground between these and top-down structured approaches such as Bayesian cognitive modeling has emerged: That models posed in different description languages may *not* be at odds simply because they are posed at different levels of analysis, and in fact *should* be tested for complementarity (Rogers & McClelland, 2008; Griffiths, Vul, & Sanborn, 2012).

It is this integrative approach that I view as the most fruitful in examining the validity of meta-learning and meta-learned cognitive models precisely because (1) it allows us to address the challenges of working within a single paradigm (say, the lack of interpretability of a connectionist approach) at the same time as (2) providing stronger grounds on which to refute a cognitive model (say, by its inconsistency with evidence from neural recordings, or its inability to account for how an ecological task is solved). Making use of the former benefit is especially critical, as the meta-learned models commonly employed, including by Binz et al., have the potential to be even more inscrutable than a connectionist model initialized in a data-agnostic way.

Binz et al. discuss two studies of meta-learning and meta-learned models that bridge levels of analysis in this manner: Firstly, a meta-learning algorithm has been tested against experimental neuroscience findings in prefrontal cortex (Wang et al., 2018). Secondly, a meta-learned recurrent neural network can approximate the posterior predictive distribution picked out as optimal by a Bayesian approach (Ortega et al., 2019). Connecting neuroscientific findings with computational-level analysis via algorithm is an exciting result. However, as Binz et al. note, the goodness of fit of the meta-learned approximation employed in both studies is not guaranteed, and has been empirically demonstrated to be poor.

As a contrast to an approach that makes use of approximation, our work (Grant, Finn, Levine, Darrell, & Griffiths, 2018) draws a formal connection between a connectionist implementation of meta-learning and inference in a hierarchical Bayesian model by making precise the prior, likelihood, and parameter estimation procedure implied in the use of the meta-learning implementation. Equivalently, this result describes a way to implement a rational solution to a problem of learning-to-learn in a connectionist architecture (though there are likely to be many equivalent implementations). A formal integration across levels like this is tighter than an approximation approach, and therefore provides a firmer footing for integrative constraints across levels of analysis.

Follow-up investigations have made use of this connection between computational-level and algorithmic-level approaches. For example, in McCoy, Grant, Smolensky, Griffiths, and Linzen (2020), we used an analogous setup to Grant et al. (2018) to meta-learn a syllable typology in a limited data setting akin to an impoverished language learning environment. To better accommodate the complex dynamics of learning, we relaxed some constraints on the meta-learning algorithm, thus for the moment doing away with the tight connection between the algorithmic and computational levels. However, in sticking with methods – namely tuning the gradient-based initialization for learning in a neural network – for which ongoing research in machine learning is formally characterizing how prior knowledge (Dominé, Braun, Fitzgerald, & Saxe, 2023), including data-driven prior knowledge (Lindsey & Lippl, 2023), interacts with the learning algorithm and environment, my view is that these approaches will soon benefit from tighter connections between the algorithmic and computational levels echoing to the connection derived in Grant et al. (2018).

Absent these connections, because meta-learning and meta-learned models are underconstrained and data-driven, it is challenging to evaluate the validity and implications of these

models for our understanding of how experience shapes learning. Thus, scientists interested in the place of meta-learning and meta-learned models in cognitive science should work to make precise the constraints that these models imply across levels of analysis, including by making use of analytical techniques from machine learning, at the same time looking into complementary constraints from experimental neuroscience, and ecologically relevant environments. Given that so many aspects remain open, it is an exciting time to be working with and on meta-learning toolkit.

Acknowledgments. N/A.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Dominé, C. C., Braun, L., Fitzgerald, J. E., & Saxe, A. M. (2023). Exact learning dynamics of deep linear networks with prior knowledge. *Journal of Statistical Mechanics: Theory and Experiment*, *2023*(11), 114004. https://doi.org/10.1088/1742-5468/ad01b8

Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, *21*(4), 263–268. https://doi.org/10.1177/0963721412447619

Lindsey, J. W., & Lippl, S. (2023). Implicit regularization of multi-task learning and fine-tuning in overparameterized neural networks. arXiv preprint arXiv:2310.02396. https://doi.org/10.48550/arXiv.2310.02396

McCoy, R. T., Grant, E., Smolensky, P., Griffiths, T. L., & Linzen, T. (2020). Universal linguistic inductive biases via meta-learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. https://doi.org/10.48550/arXiv.2006.16324

Ortega, P. A., Wang, J. X., Rowland, M., Genewein, T., Kurth-Nelson, Z., Pascanu, R., … Legg, S. (2019). Meta-learning of sequential strategies. arXiv preprint arXiv:1905.03030. https://doi.org/10.48550/arXiv.1905.03030

Rogers, T. T., & McClelland, J. L. (2008). A simple model from a powerful framework that spans levels of analysis. *Behavioral and Brain Sciences*, *31*(6), 729–749. doi:10.1017/S0140525X08006067

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*(1), 1–23. https://doi.org/10.1017/S0140525X00052432

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., … Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience 21*, 860–868. https://doi.org/10.1038/s41593-018-0147-8

Learning and memory are inextricable

Sue Llewellyn*

University of Manchester, Manchester, UK
sue.llewellyn@manchester.ac.uk
https://www.humanities.manchester.ac.uk/

*Corresponding author.

doi:10.1017/S0140525X2400013X, e153

Abstract

The authors' aim is to build "more biologically plausible learning algorithms" that work in naturalistic environments. Given that, first, human learning and memory are inextricable, and, second, that much human learning is unconscious, can the authors' first research question of how people improve their learning abilities over time be answered without addressing these two issues? I argue that it cannot.

Learning is the process of acquiring a memory; learning and memory both depend, fundamentally, on association (Fuster, 1999). The inextricability of learning and memory originates because any to-be-learned information is encoded in memory through association, the same associative encoding also enables retrieval from memory (Brown & Craik, 2000; Tulving & Thomson, 1973). Once retrieved, the encoded memory will drive expectations in the same (or similar) environments to those in which the learning took place.

The authors identify their meta-learned models as ones that "acquire their inductive biases from experience, i.e. by repeatedly interacting with an environment." This implies that learning is cumulative over time as the model repeatedly samples the environment. For human learning to be cumulative across time, the learned information must be encoded and retained in memory. Later in the paper, the authors, briefly, acknowledge the contribution of episodic memory and the hippocampus but do not spell out how hippocampal memory systems impact their models.

In relation to the authors' first research question, people can improve their learning abilities through *elaborate* encoding which creates associations between the to-be-learnt episodic material and information already encoded in episodic memory networks (Foer, 2011; Llewellyn, 2013; Yates, 1966). The hippocampus is crucial to both associative encoding and later retrieval of episodes through association (Davachi & Wagner, 2002; Llewellyn, 2013; Thakral, Benoit, & Schacter, 2017). The sequential, associative nature of elements of learned, encoded and retained episodic memories gives rise to rational, step-by-step learning. However, during rapid eye movement (REM) sleep, the hippocampus may take *elements of different* episodic experiences to identify more elaborative, associative, probabilistic and meaningful patterns in experience which may be expressed, visually, as dreams and instantiated as cortical nodes/junctions during non-rapid eye movement (NREM) sleep (Llewellyn & Hobson, 2015). Cortical episodic memory networks then consist of sequential, associative pathways which converge at nodes/junctions which express patterned, probabilistic, experiential learning. The hidden units/nodes in neural networks may be analogous to these nodes/junctions in cortical episodic memory networks.

Given the authors' recognition, "that episodic memory could be the key to explaining human performance in naturalistic environments" it seems pertinent to expand their algorithms to encompass not only the logical, experiential, "step-by-step" learning patterns that predominate during wake but also the more complex, hyper-associative, experiential patterns identified during REM sleep. Admittedly, this extension may be, somewhat, futuristic. But research has already mimicked slow wave sleep which restored stability to neural networks engaged in unsupervised learning (Watkins, Kim, Sornborger, & Kenyon, 2020). Also, given the adaptivity, speed via massive parallelism, fault tolerance and optimality of neural networks (Bezdek, 1992) and that machine learning programmes trained on probabilistic reasoning are superior to the human brain for visual pattern recognition (Pavlus, 2016) it may be possible that, in the recurrent neural networks mobilized by the authors, REM sleep is not required for complex pattern recognition.

Experiences in natural environments may never be re-enacted in their entirety but they are certainly non-random and do not exclude expectations. Across evolutionary time, many interactions with the environment held dangers because of predators and

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

competitors. To try to avoid dangers, humans needed to identify the probabilistic, behavioural patterns of predators and competitors. Such patterns would only be revealed over several episodes, as humans observed the associations that drove predator behaviour. For example, lions tend to visit waterholes at night when prey are abundant (lion presence is associated with nighttime) but, in the dry season, lions get so thirsty that they may be at the waterhole during the day (in the dry season, lion presence can be associated with daytime). Also, elephants tend to drive lions away from the waterhole, so the presence of elephants offers some safety (elephant presence is associated with lion absence). When many associations are at stake and/or actions must be fast, probabilistic associative patterns, derived from multiple episodes, drive expectations and learning during future interactions in the environment.

Much, probably the majority, of learning occurs unconsciously. Contemporarily, dangers may differ but unconscious expectations, formed through learning and retained as unconscious memories, still drive fast responses to threats (Öhman, Carlsson, Lundqvist, & Ingvar, 2007). Concomitantly, it would be expected that the elaborative, complex, associative, probabilistic and meaningful experiential patterns formed in REM sleep would be unconscious. Arzi et al. (2012) showed that new associations, learned during REM sleep, were retained as unconscious memories, then functioning, during wake, as unconscious expectations which drove actions.

Does the human unconscious have any parallels in machine learning? We know that machine learning algorithms inherit unconscious human biases, indeed they can amplify them (Thomas, 2018). Moreover, humans can continue to reproduce machine learning bias, even after they no longer use the algorithm (Vicente & Matute, 2023). Clearly, unconscious learning and memory are not subject to voluntary control. Although unconscious biases can be detrimental (e.g., to marginalized groups) inductive, unconscious (or implicit) biases are essential for faster information processing both in humans and machine learning. Inductive biases depend on the associations formed through experience, in machine learning such associations occur during unsupervised learning. Specifically, in the authors' meta-learned models, associations will be experiential.

In sum, the authors assert that their meta-learned models are "invaluable tools to study, analyse and understand the human mind." In humans, experiential, associative learning becomes associative memories that, then, drive expectations in subsequent learning. In the memory network architecture of the brain/mind these associative memories may take the form of the serial, associative pathways which underlie conscious learning. Equally, where these serial pathways cross at nodes/junctions, complex, hyper-associative, experiential patterns arise from elements of the different intersecting pathways. These patterns are retained as unconscious memories which associate elements of different experiences.

Association is fundamental to both learning and memory, whether conscious or unconscious. The authors' first research question of how people improve their learning abilities over time can be better addressed through acknowledging, first, the inextricability of learning and memory and, second, the role of conscious and unconscious association in each.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interests. None.

References

Arzi, A., Shedlesky, L., Ben-Shaul, M., Nasser, K., Oksenberg, A., Hairston, I. S., & Sobel, N. (2012). Humans can learn new information during sleep. *Nature neuroscience*, *15* (10), 1460–1465. doi: 10.1038/nn.3193

Bezdek, J. C. (1992). On the relationship between neural networks, pattern recognition and intelligence. *International Journal of Approximate Reasoning*, *6*(2), 85–107.

Brown, S. C., & Craik, F. I. M. (2000). Encoding and retrieval of information. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 93–107). Oxford University Press.

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: Insights from relational and item-based learning. *Journal of Neurophysiology*, *88*(2), 982–990. doi: 10.1152/jn.2002.88.2.982

Foer, J. (2011). *Moonwalking with Einstein: The art and science of remembering everything*. Allen Lane.

Fuster, J. M. (1999). *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. MIT Press.

Llewellyn, S. (2013). Such stuff as dreams are made on? Elaborative encoding, the ancient art of memory, and the hippocampus. *Behavioral and Brain Sciences*, *36*(6), 589–607. doi: 10.1017/S0140525X12003135

Llewellyn, S., & Hobson, J. A. (2015). Not only… but also: REM sleep creates and NREM stage 2 instantiates landmark junctions in cortical memory networks. *Neurobiology of Learning and Memory*, *122*, 69–87. doi: 10.1016/j.nlm.2015.04.005

Öhman, A., Carlsson, K., Lundqvist, D., & Ingvar, M. (2007). On the unconscious subcortical origin of human fear. *Physiology & behavior*, *92*(1–2), 180–185. doi: 10.1016/j.physbeh.2007.05.057

Pavlus, J. (2016). Computers now recognize patterns better than humans can. *Scientific American*. Originally published as, "8. Sight-Reading Software" in Scientific American Magazine Vol. 315 No. 6 (December 2016), p. 41. doi: 10.1038/scientificamerican1216-39b

Thakral, P. P., Benoit, R. G., & Schacter, D. L. (2017). Characterizing the role of the hippocampus during episodic simulation and encoding. *Hippocampus*, *27*(12), 1275–1284. doi: 10.1002/hipo.22796

Thomas, R. (2018). Analyzing & preventing unconscious bias in machine learning. Retrieved from: QCon.ai 2018.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373.

Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports*, *13*(1), 15737. doi: 10.1038/s41598-023-42384-8

Watkins, Y., Kim, E., Sornborger, A., & Kenyon, G. T. (2020). Using sinusoidally-modulated noise as a surrogate for slow-wave sleep to accomplish stable unsupervised dictionary learning in a spike-based sparse coding model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 360–361).

Yates, F. A. (1966). *The art of memory*. Routledge and Kegan Paul.

Meta-learning: Bayesian or quantum?

Antonio Mastrogiorgio*

Department of Psychological and Social Sciences, John Cabot University, Rome, Italy
www.johncabot.edu
mastrogiorgio.antonio@gmail.com
https://sites.google.com/site/mastrogiorgioantonio/

*Corresponding author.

doi:10.1017/S0140525X24000220, e154

Abstract

Abundant experimental evidence illustrates violations of Bayesian models across various cognitive processes. Quantum cognition capitalizes on the limitations of Bayesian models, providing a compelling alternative. We suggest that a generalized quantum approach in meta-learning is simultaneously more robust and flexible, as it retains all the advantages of the Bayesian framework while avoiding its limitations.

The use of the Bayesian framework in meta-learning represents an elegant way to bypass the strictures of an ex-ante specification of models of cognition. As proposed by Binz et al. (hereafter the Authors), Bayesian inference, building upon unconstrained interactions with the environment, represents a viable alternative to more traditional hand-designed learning algorithms.

However, Bayesian models come with inherent limitations, which researchers are rarely aware of. While scholars normally endorse Bayesian models because of their unconstrained features, they rarely consider that such models are actually "constrained" to the Kolmogorovian assumptions of classical probability theory.

Abundant experimental evidence illustrates violations of Bayesian models across various cognitive processes, including probability judgment errors, memory recognition, semantic spaces, information processing, learning, concept combination, and perception (e.g., Pothos and Busemeyer, 2022). These violations stem from the fact that many cognitive phenomena do not adhere to the law of total probability, along with the distributivity axiom, assumed in classical Kolmogorovian probability. Consequently, they do not admit a Bayesian operationalization.

Let us consider a stylized meta-learning process (similar to those discussed by the Authors) that can be operationalized through a Bayesian model. Suppose we hypothesize that learning performance can assume the (mutually exclusive) states P1 or P2, conditioned on meta-experience, which can assume the (mutually exclusive) states E1 or E2. Let's aim to predict the total probability, p, of the state P2. Consistent with a Bayesian framework, we can consider two cases: One, which we'll refer to as the "unconditioned case," where we only observe p(P2); and the other, referred to as the "conditioned case," where we observe the conditioned probabilities, p(P2|E1) and p(P2|E2).

Now, suppose that in experimental settings, the observed data reveal that in the "unconditioned case," the probability is p(P2) = 0.29, while in the "conditioned case," the probability is p(E1)·p(P2|E1) + p(E2)·p(P2|E2) = 0.59, where p(P2|E2) = 0.63.

We quickly realize that these experimental results are incompatible with a Bayesian framework: The total probability of P2 in the "unconditioned case" is inconsistent with that of the "conditioned case" (0.29 vs. 0.59), and the total probability of the "unconditioned case" is also lower than that in the "conditioned case," restricted to E2 (0.29 vs. 0.63). When evidence violates the law of total probability, Bayesian models reveal inadequacies (for further discussion, see Busemeyer & Wang, 2015; Busemeyer, Wang, & Lambert-Mogiliansky, 2009).

Such situations are paradoxical: They are somehow implausible to the extent that they do not admit a Bayesian formalization, yet they result from experimental evidence. When it comes to understanding such evidence, quantum cognition comes to the fore as a burgeoning field of research that dialectically capitalizes on the limitations of Bayesian models, providing a compelling alternative. A fundamental difference between Bayesian and quantum models – relevant for the present critique – lies in the fact that quantum models can account for such evidence precisely because they do not adhere to the law of total probability (for a comprehensive comparison between Bayesian and quantum models in cognition, refer to Bruza, Wang, & Busemeyer, 2015).

Continuing with the example discussed above, what differs between the "unconditioned case" and "conditioned case" is that, respectively, the non-observation or the observation of the conditioning variable (E) is not neutral for the final probability. In other words, the two models are incompatible and do not admit a mutually consistent formalization.

In a quantum framework, the violations of the total law of probability are due to *interference effects*, which occur when the conditioning variables are not observed. In our example, the interference between the two conditions (E1 and E2) implies that they behave like waves, where the interference can be either destructive (canceling) or constructive (resonating), affecting the final probability p(P2). On the contrary, when E is observed, the result is compatible with a classical framework as the act of measuring the mutually exclusive states of experience eliminates their interference (for an overview on the role of interference effects, see Busemeyer & Bruza, 2012).

The Authors wisely avoid claiming that meta-learning is the ultimate solution to every modeling problem, and they contemplate, in what they call "Intricate training processes," the possibility that "the resulting model [of meta-learning] does not fit the observed data."

We think that such situations are not just due, as implicitly suggested by the Authors, to the complexity of the scenarios, but to the fact that the tacit probabilistic assumptions of Bayesian models are somehow too restrictive.

More in detail, the plausibility of neurocognitive Bayesian foundations of meta-learning would require stronger justifications. Indeed, assuming that prefrontal circuits may constitute a meta-reinforcement learning system (cf., Wang et al., 2018) or, in general terms, that the brain is a Bayesian machine, matching top-down prediction with bottom-up experience (cf., Friston, 2010), would also imply assuming that Kolmogorovian probability ascribes to the biological realm.

The employment of Bayesian models is institutionalized to such an extent that their foundational assumptions are rarely contested in scientific debates. However, there are situations, like the one discussed here, in which Bayesian models reveal their limitations: Despite aspiring to provide a generalized framework for meta-learning, they inherently harbor very restrictive assumptions in probability theory.

On the contrary, quantum models exhibit greater flexibility, are more robust, and can offer a more sophisticated view of the neurocognitive mechanisms involved in human learning (cf., Mastrogiorgio, 2022).

However, quantum cognition is not a tout court alternative to Bayesian models but rather a generalization applicable to cases where the law of total probability is violated. This implies that Bayesian models represent a special case within the broader quantum framework: Quantum models reduce to Bayesian models when experimental evidence aligns with the requirements of the distributivity axiom and the law of total probability.

Precisely because we support the Authors' proposal of employing unconstrained logics in meta-learning, we also believe that a generalized quantum approach is simultaneously more robust and flexible, as it retains all the advantages of the Bayesian framework while avoiding its limitations.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015). Quantum cognition: A new theoretical approach to psychology. *Trends in Cognitive Sciences*, *19*, 383–393.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge University Press.

Busemeyer, J. R., & Wang, Z. (2015). What is quantum cognition, and how is it applied to psychology?. *Current Directions in Psychological Science*, *24*(3), 163–169.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). Empirical comparison of Markov and quantum models of decision making. *Journal of Mathematical Psychology*, *53*, 423–433.
Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.
Mastrogiorgio, A. (2022). A quantum predictive brain: Complementarity between Top-down predictions and bottom-up evidence. *Frontiers in Psychology*, *13*, 869894.
Pothos, E.M., & Busemeyer, J.R. (2022). Quantum cognition. *Annual Review of Psychology*, *73*, 749–778.
Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z.,… Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, *21*(6), 860–868.

Meta-learning as a bridge between neural networks and symbolic Bayesian models

R. Thomas McCoy[a] and Thomas L. Griffiths[b]*

[a]Department of Linguistics, Yale University, New Haven, CT, USA and [b]Departments of Psychology and Computer Science, Princeton University, Princeton, NJ, USA
tom.mccoy@yale.edu
tomg@princeton.edu
https://rtmccoy.com/
http://cocosci.princeton.edu/tom/

*Corresponding author.

doi:10.1017/S0140525X24000116, e155

Abstract

Meta-learning is even more broadly relevant to the study of inductive biases than Binz et al. suggest: Its implications go beyond the extensions to rational analysis that they discuss. One noteworthy example is that meta-learning can act as a bridge between the vector representations of neural networks and the symbolic hypothesis spaces used in many Bayesian models.

Like many aspects of cognition, learning can be analyzed at multiple levels. At a high level (Marr's [1982] "computational" level) we can model learning by providing an abstract characterization of the learner's inductive biases: The preferences that the learner has for some types of generalizations over others (Mitchell, 1997). At a lower level, learning can be modeled by specifying the particular algorithms and representations that the learner uses to realize its inductive biases. For each of these levels, there are modeling traditions that have been successful: Rational analysis and Bayesian models are defined at the computational level, while neural networks are defined at the level of algorithm and representation. But how can we connect these different traditions? How can we work toward unified theories that bridge the divide between levels? In this piece, we agree with, and extend, Binz et al.'s point that meta-learning is a powerful tool for studying inductive biases in a way that spans levels of analysis.

Binz et al. describe how an agent can use meta-learning to derive inductive biases from its environment. This makes meta-learning well-suited for modeling situations where human inductive biases align with some problem that humans face – the situations that are well-covered by the paradigm of rational analysis (Anderson, 1990). As Binz et al. discuss, meta-learning can therefore be used to enable an algorithmically defined model (such as a neural network) to find the solution predicted by rational analysis, a procedure that bridges the divide between abstract rational solutions and specific algorithmic instantiations.

This direction laid out by Binz et al. is exciting. We argue that it can in fact be viewed as one special case within a broader space of possible lines of inquiry about inductive biases that meta-learning opens up. In the more general case, the Bayesian perspective allows us to define an inductive bias as a probability distribution over hypotheses. A neural network can meta-learn from data sampled from this distribution, giving it the inductive bias in question. The distribution that is used could be drawn from (an approximation of) a human's experience, in which case this framing matches the extension of rational analysis that Binz et al. advocate for. But it is also possible to use other approaches for defining this distribution, which can correspond to any probabilistic model. Since we can control probabilistic models, using a probabilistic model to define the distribution makes it possible to control the inductive biases that the meta-learned model ends up with (Lake, 2019; Lake & Baroni, 2023; McCoy, Grant, Smolensky, Griffiths, & Linzen, 2020). This allows us to take an inductive bias defined at Marr's computational level and distill it into a neural network defined at the level of algorithm and representation.

Traditionally, certain types of inductive biases have been associated with certain types of algorithms and representations: The strong inductive biases of Bayesian models have generally been based on discrete, symbolic representations (e.g., Goodman, Tenenbaum, Feldman, & Griffiths, 2008), while neural networks use continuous vector representations (Hinton, McClelland, & Rumelhart, 1986) and have weak inductive biases. However, meta-learning enables us to separately manipulate inductive biases and representations, making it possible to model previously inaccessible combinations of representations and inductive biases. One noteworthy example is that we can use meta-learning to give symbolic inductive biases to a neural network, allowing us to study whether and how structured hypothesis spaces (of the sort often used in Bayesian models) can be realized in a system with continuous vector representations (the type of representation that is central in both biological and artificial neural networks). Thus, while Binz et al. note that meta-learning can be used as an alternative to Bayesian models, another use of meta-learning is in fact to expand the applicability of Bayesian approaches by reconciling them with connectionist models – thereby bringing together two successful research traditions that have often been framed as antagonistic (e.g., Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; McClelland et al., 2010).

In our prior work, we have demonstrated the efficacy of this approach in the domain of language (McCoy & Griffiths, 2023). We started with a Bayesian model created by Yang and Piantadosi (2022), whose inductive bias is defined using a symbolic grammar. We then used meta-learning (specially, MAML: Finn, Abbeel, & Levine, 2017; Grant, Finn, Levine, Darrell, & Griffiths, 2018) to distill this Bayesian model's prior into a neural network. The resulting system had strong inductive biases of the sort traditionally found only in symbolic models, enabling this system to learn formal linguistic patterns from small numbers of examples despite being a neural network, a class of systems that normally requires far more examples to learn such patterns. Additionally, the flexible neural implementation of this system

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

made it possible to train it on naturalistic textual data, something that is intractable with the Bayesian model that we built on. Thus, meta-learning enabled the creation of a model that combined the complementary strengths of Bayesian and connectionist models of language learning.

These results show that inductive biases traditionally defined using symbolic Bayesian models can instead be realized inside a neural network. Therefore, symbolic inductive biases do not necessarily require inherently symbolic representations or algorithms. This demonstration provides one already-realized example of how meta-learning can advance our understanding of foundational questions about how different levels of cognition relate to each other, in ways that go beyond the realm of rational analysis.

Financial support. This material is based upon work supported by the National Science Foundation SBE Postdoctoral Research Fellowship under Grant No. 2204152 and the Office of Naval Research under Grant No. N00014-18-1-2873.

Competing interests. None.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.
Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, 1126–1135.
Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.
Grant, E., Finn, C., Levine, S., Darrell, T., & Griffiths, T. (2018). Recasting gradient-based meta-learning as hierarchical Bayes. *International Conference on Learning Representations*.
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.
Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. Foundations, pp. 77–109.
Lake, B. M. (2019). Compositional generalization through meta sequence-to-sequence learning. *Advances in Neural Information Processing Systems*, *32*.
Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*(7985), 115–121.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman.
McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356.
McCoy, R. T., Grant, E., Smolensky, P., Griffiths, T. L., & Linzen, T. (2020). Universal linguistic inductive biases via meta-learning. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, 737–743.
McCoy, R. T., & Griffiths, T. L. (2023). Modeling rapid language learning by distilling Bayesian priors into artificial neural networks. *arXiv preprint arXiv:2305.14701*.
Mitchell, T. M. (1997). *Machine learning*. McGraw Hill.
Yang, Y., & Piantadosi, S. T. (2022). One model for the learning of language. *Proceedings of the National Academy of Sciences*, *119*(5), e2021865119.

Meta-learned models beyond and beneath the cognitive

Mihnea Moldoveanu*

Desautels Centre for Integrative Thinking, Rotman School of Management, University of Toronto, Toronto, ON, Canada
mihnea.moldoveanu@rotman.utoronto.ca

*Corresponding author.

doi:10.1017/S0140525X24000141, e156

Abstract

I propose that meta-learned models, and in particular the situation-aware deployment of "learning-to-infer" modules can be advantageously extended to domains commonly thought to lie outside the cognitive, such as motivations and preferences on one hand, and the effectuation of micro- and coping-type behaviors.

An account of the ways in which locally meta-learned models can address difficulties arising from computational intractability of Bayesian inference in non-small worlds and difficulties in articulating inference problems over unknown state spaces have is overdue. We have for some time been aware that the experimental evidence base of behavioral decision theory admits of alternative plausible explanations via models of the experimental condition (meant to illustrate "base rate neglect," or availability of a plausible simile) that explain how the subject can seem biased to an observer while in fact engaging in a reasonable and ecologically successful pattern of inference (Moldoveanu & Langer, 2002; Marsh, Todd, & Gigerenzer, 2004). At the same time, incorporating the informational (say, "bits") and computational ("operations performed upon bits") costs of inferential calculations in models of cognition not only "makes sense" – as storage and calculation both require work – but can rationalize patterns of behavior that previously appeared to some as irrational or sub-rational (Gershman, Horvitz, & Tennenbaum, 2015; Moldoveanu, 2011).

The meta-learning program of inquiry can be extended to phenomena and episodes that lie beyond or on the fringes of what we would call "cognitive," at both the "upstream" (motives, motivations, identities) and "downstream" (motor behaviors, perceptual inferences) ends. Motivations, "preferences," and meta-preferences can be understood as the outcomes of a process by which one learns from one's own reactions to an environment's responses to one's own behaviors. At the other end, in-the-moment "heedful coping" for the purpose of getting a "maximal grip" (Merleau Ponty, 2012[1974]) on an immediate physical or interpersonal environment can be described in ways that make transparent the structure of the inferential learning problem and the function that one is trying to approximate, for example: Producing situation-specific, successful combinations of vertical and horizontal pressures through one's digits in order to balance a large, flat object, or of activations of muscles to produce particular facial and postural expressions meant to cause someone else to do, say or think in a particular way.

Consider, first, the "motivational" nexus of motivation-identity-preference. The idea that one infers one's own "attitudes" from one's behavior in situations whose affordances encode "choicefulness" has been around for a while (Bem, 1967), but the problem of "learning to prefer (X to Y, say)" is neither immediately self-evident or "given" nor, once posed, computationally simple (or, even tractable). But, once preferences are understood as dispositions to act in particular ways or combinations of ways in a context and encoded by conditional probabilities linking combinations of sensorial fields and internal states to actions, the self-referential problem of inferring "What do I like?/What motivates me?" from observations of "what do I do (or, choose to do) in situations in which…?" can get off the ground. One can develop preferences that are highly specific (combinations of ingredients mixed in precise sequences and proportions to make a "dish") or general ("I prefer injustice to disorder") and motivations that are domain-specific ("decentralized decision rights as an approach to managing *this* team on *this* task") or less so ("to

enable and facilitate other people's sense of autonomy"). Preferences and motivations can be more or less sophisticated (in terms of the number of features of a "situation" the inference of preference depends on, and the logical depth or computational complexity of the inference algorithm) and more or less context-adaptive, and more or less susceptible to recursive refinement upon experiencing the ways in which one behaves in different environments. Thus, being able to adaptively modify the informational and computational complexity of the learning algorithm adds a much-needed degree of freedom to the modelers toolkit. Both preferences and motivations may be learned without an explicit awareness of that which is learned or that which learning is conditioned by. "Learning to prefer" (or learning to be the self one, motivationally, is) can thus be phrased in a way that tracks "learning to infer", provided we can successfully formulate a local objective or cost function the inferential process meliorates or seeks to extremize. "Internal conflicts" appear as neither pathological nor irrational: An optimal (or ecologically adapted in virtue of its adaptiveness) brain can comprise a set of independent agents that have conflicting objectives (Livnat & Pippenger, 2006) but are jointly adaptive to environmental niches its organism copes with frequently enough.

Second, consider the "in-the-moment" perception-sensation-effectuation nexus, which has to do with trying to get a target variable to take on a certain value within a certain time window using the least or a fixed amount of energy – such as controlling the vertical displacement of a tray of containers containing hot liquids, the horizontal displacement of an inverted pendulum, the angular velocity, height, and vertical velocity of a coin used in a coin toss or even a dynamical network of physiognomical micro-responses in an emotionally charged meeting with several attendees. In such cases, the inverse, counterfactual-dependent problem of "causal inference" is replaced by the forward, direct problem of effectuation (or control): One is attempting to produce and in some cases also maintain a specific set of values of variables X_T (vertical displacement of a tray of liquids, perceived visceral or emotional responses of another person) that form part of the state space $\boldsymbol{X}$ of a dynamically evolving system $\boldsymbol{X_t = F(X,U,\ t)}$ by making specific changes to one or more "input" or lever variables $\boldsymbol{U}$ on the basis of observations of some proxy variables $\boldsymbol{Y}$ that encode or register filtered, biased states of $\boldsymbol{X}$ via $\boldsymbol{Y = G(X,t)}$. In this case, the underlying inference problem can be posed in terms of maximizing the time-bounded and energy-efficient controllability and observability of $\boldsymbol{X_t = F(X,U,\ t);\ Y = G(X,t)}$ for instance by the optimal choice and placement of the "lever" or "driver" nodes (Li and Laszlo Barabassi, 2016), or in terms of making changes to the structural properties of $\boldsymbol{X_t = F(X,U,\ t);\ Y = G(X,t)}$ in ways that alter its temporal dynamics or time constants (for instance, learning to control a tremor by using different combinations or muscle groups for effecting a fine movement, which changes the parameters of $\boldsymbol{X_t = F(X,U,\ t);\ Y = G(X,t)}$ and thus the pole-zero distribution of its transfer function).

Financial support. This research was funded by the Desautels Centre for Integrative Thinking, Rotman School of Management, University of Toronto.

Competing interest. None.

References

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*, 183–200.

Gershman, S. J., Horvitz, E. J., & Tennenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds and machines. *Science*, *349* (6245), 273–278.

Liu, Y. -Y., & Laszlo-Barabassi, A. (2016). Control principles of complex systems. *Review of Modern Physics*, *88*(3), 1–58.

Livnat, A., & Pippenger, N. (2006). An optimal brain can be composed of conflicting agents. *Proceedings of the National Academy of Sciences*, *103*(9), 3198–3202.

Marsh, B., Todd P. M., & Gigerenzer, G. (2004). Cognitive heuristics: Reasoning the fast and frugal way. In J. P. Leighton, & R. J. Sternberg (Eds.), *The nature of reasoning* (pp. 273–287). Cambridge University Press.

Merleau Ponty, M. (2012)(1974). *The phenomenology of perception*. Routledge.

Moldoveanu, M. C. (2011). *Inside man: The discipline of modeling human ways of being*. Stanford University Press.

Moldoveanu, M. C., & Langer, E. J. (2002). False memories of the future: A critique of the applications of probabilistic reasoning to the study of cognitive processes. *Psychological Review*, *109*(2), 358–375.

Meta-learned models as tools to test theories of cognitive development

Kate Nussenbaum[a]* and Catherine A. Hartley[b]*

[a]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA and [b]Department of Psychology, New York University, New York, NY, USA
katenuss@princeton.edu
cate@nyu.edu
https://www.katenuss.com/
https://www.hartleylab.org/

*Corresponding authors.

doi:10.1017/S0140525X24000281, e157

Abstract

Binz et al. argue that meta-learned models are essential tools for understanding adult cognition. Here, we propose that these models are particularly useful for testing hypotheses about why learning processes change across development. By leveraging their ability to discover optimal algorithms and account for capacity limitations, researchers can use these models to test competing theories of developmental change in learning.

Binz et al. argue compellingly for meta-learning as a tool to understand adult cognition, but their vision is incomplete: Meta-learned models are particularly apt tools for studying development. As the authors note, meta-learned models provide a natural foundation for theorizing about learning to learn (Wang, 2021) and for understanding experience-driven changes in goal-directed behavior (Nussenbaum & Hartley, in press). Here, however, we consider the authors' central argument and focus not on meta-learning algorithms or the process of "learning to learn," but rather on meta-learned models as a *tool* for understanding developmental changes in learning. To date, research has revealed age-related shifts in the algorithms and neural circuitry that underlie learning (Bolenz, Reiter, & Eppinger, 2017; Gualtieri & Finn, 2022; Hartley, Nussenbaum, & Cohen, 2021; Nussenbaum & Hartley, 2019; Raab & Hartley, 2018), but understanding why these shifts occur has proven more difficult.

Disentangling whether age-related changes in learning reflect adaptation to age-varying "external" ecological problems *or* "internal" changes in cognitive capacity has been difficult, in part because developmentalists have lacked the formalizations needed to make specific predictions about how these two factors might drive age-related changes in behavior. As an example, in

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

many studies of reinforcement learning, participants are tasked with earning the most reward (e.g., points or money) by selecting between different options. In these tasks, children tend to make "noisier" or more exploratory choices, which typically result in poorer performance (Nussenbaum & Hartley, 2019) but enhanced learning of task structure (Blanco & Sloutsky, 2021; Liquin & Gopnik, 2022; Sumner et al., 2019). These findings have led researchers to theorize that children and adults have optimized their learning computations to maximize reward over different types of environments – children's learning contexts may have features (e.g., greater reward stochasticity, longer temporal horizons) that favor exploration over immediate reward gain (Gopnik, 2020). An alternative account, however, is that children's "noisier" behavior reflects a more limited cognitive capacity (Craik & Bialystok, 2006; Ruel, Devine, & Eppinger, 2021), which constrains their "optimization" of learning computations.

By enabling separable manipulation of external experience and internal cognitive capacity, meta-learned models may help arbitrate between these accounts of developmental change. As Binz et al. note, through meta-learning, the optimal algorithm emerges through experience, and is shaped by the distribution of environments on which the model is trained. Thus, these models can be used to test how differences in "training" experience might yield different patterns of learning at "test." Examining how training environments influence model behavior within the same tasks that have been used to study cognitive development can provide insight into the types of environments for which children, adolescents, and adults have optimized their learning computations. As one example, Binz and Schulz (2022) found that increasing the variance in the rewards that a model experienced during training led to exploration decisions at test that more closely approximated those of human learners, suggesting that people may have tuned their learning computations for environments with more stochasticity than the task context. Importantly, meta-learned models can provide insight into how optimal learning is shaped by exposure to highly complex environments in which multiple features vary and interact – environments for which analytically derived "solutions" are intractable but that are more reflective of real-world contexts than the simple tasks that have been used in most prior research.

At the same time, meta-learned models can account for how changes in capacity limitations may yield developmental changes in learning. As Binz et al. note, the complexity of meta-learned models can be manipulated easily, particularly when they are implemented as neural networks. Manipulating capacity constraints and comparing model behavior to that of learners at different ages can thus provide insight into whether developmental changes in learning are well accounted for by theories of "resource-rationality." Binz and Schulz (2022) found, for example, that increasing the complexity of the algorithms that a meta-learning network could implement led to changes in patterns of directed exploration that mirrored those that occur across adolescence (Somerville et al., 2017). This result exemplifies how, rather than emerging from differences in external "training" experience, age-related changes in learning can also be driven by differences in cognitive capacity.

Further, neural network implementations of meta-learning enable separable manipulations of "algorithmic complexity" via network weights and "computational complexity" via network activations. Future developmental modeling work could explore the ramifications of constraints on algorithmic *versus* computational complexity. Numerous studies have revealed a dissociation between children's knowledge of rules or structure and their ability to leverage them to guide behavior (Decker, Otto, Daw, & Hartley, 2016; Zelazo, Frye, & Rapus, 1996). The types of structural knowledge that can be acquired at different ages may depend on algorithmic complexity, while use of that knowledge may rely on processes like working memory and proactive cognitive control, which may instantiated via complex computations. Behavioral dissociations between algorithmic and computational complexity may be mirrored by neural dissociations. Age-related change in brain structure or the wiring of neural circuitry may be analogous to age-related change in network weights and relate more strongly to algorithmic complexity, whereas age-related change in patterns of neural activation during learning may be more closely related to the network activations that underlie computational complexity. Thus, predictions about how age-related change in these two forms of capacity limitations affect learning could be further tested and constrained with neural data.

While we have suggested that researchers can leverage meta-learned models to more explicitly test whether children optimize their behavior for different environments *or* with different constraints, this dichotomy is likely false. Experience and constraints interact throughout the lifespan, and the changing "constraints" implemented by neurobiology may themselves serve an adaptive function – it may be the case that the complexity of the learning algorithms an organism can implement and execute systematically increases *through* exposure to increasingly varied and complex environments. Meta-learned models of cognition thus have the potential to address questions of longstanding interest in developmental science, while empirical developmental research provides a valuable testbed for the theoretical utility of these computational tools.

Acknowledgements. We thank Rheza Budiono for helpful feedback.

Financial support. Preparation of this commentary was supported by a CV Starr Foundation Fellowship (to K. N.).

Competing interest. None.

References

Binz, M., & Schulz, E. (2022). Modeling human exploration through resource-rational reinforcement learning. *Advances in Neural Information Processing Systems*, *35*, 31755–31768.

Blanco, N. J., & Sloutsky, V. M. (2021). Systematic exploration and uncertainty dominate young children's choices. *Developmental Science*, *24*(2), e13026.

Bolenz, F., Reiter, A. M. F., & Eppinger, B. (2017). Developmental changes in learning: Computational mechanisms and social influences. *Frontiers in Psychology*, *8*, 2048.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131–138.

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From creatures of habit to goal-directed learners: Tracking the developmental emergence of model-based reinforcement learning. *Psychological Science*, *27*(6), 848–858.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1803), 20190502.

Gualtieri, S., & Finn, A. S. (2022). The sweet spot: When children's developing abilities, brains, and knowledge make them better learners than adults. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *17*(5), 1322–1338.

Hartley, C. A., Nussenbaum, K., & Cohen, A. O. (2021). Interactive development of adaptive learning and memory. *Annual Review of Developmental Psychology*, *3*, 59–85.

Liquin, E. G., & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition*, *218*, 104940.

Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, *40*, 100733.

Nussenbaum, K., & Hartley, C.A. (in press). Understanding the development of reward learning through the lens of meta-learning. *Nature Reviews Psychology*.

Raab, H. A., & Hartley, C. A. (2018). The development of goal-directed decision-making. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-directed decision making: Computations and neural circuits* (pp. 279–308). Elsevier Academic Press.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Ruel, A., Devine, S., & Eppinger, B. (2021). Resource-rational approach to meta-control problems across the lifespan. *Wiley Interdisciplinary Reviews. Cognitive Science*, *12*(5), e1556.
Somerville, L. H., Sasse, S. F., Garrad, M. C., Drysdale, A. T., Abi Akar, N., Insel, C., & Wilson, R. C. (2017). Charting the expansion of strategic exploratory behavior during adolescence. *Journal of Experimental Psychology. General*, *146*(2), 155–164.
Sumner, E., Li, A. X., Perfors, A., Hayes, B., Navarro, D., & Sarnecka, B. W. (2019). The exploration advantage: Children's instinct to explore allows them to find information that adults miss. https://doi.org/10.31234/osf.io/h437v
Wang, J. X. (2021). Meta-learning in natural and artificial intelligence. *Current Opinion in Behavioral Sciences*, *38*, 90–95.
Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63.

Probabilistic programming versus meta-learning as models of cognition

Desmond C. Ong[a]*, Tan Zhi-Xuan[b], Joshua B. Tenenbaum[b] and Noah D. Goodman[c,d]

[a]Department of Psychology, University of Texas at Austin, Austin, TX, USA; [b]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA; [c]Department of Psychology, Stanford University, Stanford, CA, USA and [d]Department of Computer Science, Stanford University, Stanford, CA, USA
desmond.ong@utexas.edu
xuan@mit.edu
jbt@mit.edu
ngoodman@stanford.edu
https://cascoglab.psy.utexas.edu/desmond/
https://cocosci.mit.edu/
https://ztangent.github.io/
https://cocolab.stanford.edu/

*Corresponding author.

doi:10.1017/S0140525X24000153, e158

Abstract

We summarize the recent progress made by probabilistic programming as a unifying formalism for the probabilistic, symbolic, and data-driven aspects of human cognition. We highlight differences with meta-learning in flexibility, statistical assumptions and inferences about cogniton. We suggest that the meta-learning approach could be further strengthened by considering Connectionist *and* Bayesian approaches, rather than exclusively one or the other.

Connectionist-versus-Bayesian debates have occurred in cognitive science for decades (e.g., Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; McClelland et al., 2010), with each side progressing in theory, models, and algorithms, in turn impelling the other side to advance, resulting in a cycle of fruitful engagement. The recent summary of the meta-learning paradigm that Binz et al. proposed in the target article bridges the two by proposing how meta-learning in recurrent neural networks can address some of the traditional challenges of Bayesian approaches. But, by failing to recognize and engage with the latest iteration of Bayesian modeling approaches – including probabilistic programming as a unifying paradigm for probabilistic, symbolic, and differentiable computation (Cusumano-Towner, Saad, Lew, & Mansinghka, 2019) – this article fails to push the meta-learning paradigm as far as it could go.

The authors begin their defense of meta-learning by citing the intractability of *exact* Bayesian inference. However, this fails to address how and why meta-learning is superior to *approximate* inference for modeling cognition. As the authors themselves note, Bayesian modelers use a variety of approximate inference methods, including neural-network-powered variational inference (Dasgupta, Schulz, Tenenbaum, & Gershman, 2020; Kingma & Welling, 2013), Markov chain Monte Carlo (Ullman, Goodman, & Tenenbaum, 2012), and Sequential Monte Carlo methods (Levy, Reali, & Griffiths, 2008; Vul, Alvarez, Tenenbaum, & Black, 2009), which have all shown considerable success in modeling how humans perform inference (or fail to) in presumably intractable settings. As such, it is hardly an argument in favor of meta-learning – and against "traditional" Bayesian models – that *exact* inference is intractable.

This omission is just one way in which the article fails to engage with a modern incarnation of the Bayesian modeler's toolkit – Probabilistic Programming. In the past two decades, we have seen the development of probabilistic programming as unifying formalism for modeling the probabilistic, symbolic, and data-driven aspects of human cognition (Lake, Salakhutdinov, & Tenenbaum, 2015), as embodied in probabilistic programming language such as Church (Goodman, Mansinghka, Roy, Bonawitz, & Tenenbaum, 2012), webPPL (Goodman & Stuhlmüller, electronic), Pyro (Bingham et al., 2019), and Gen (Cusumano-Towner et al., 2019). These languages enable modelers to explore a much wider range of computational architectures than the standard meta-learning setup, which requires modelers to reformulate human cognition as a sequence prediction problem. Probabilistic programming allows modelers to unite the strengths of general-purpose predictors (i.e., neural networks) with theoretically informed constraints and model-based reasoning. For instance, Ong, Soh, Zaki, and Goodman (2021) showed how reasoning about others' emotions can be modeled by combining the constraints implied by cognitive appraisal theory with bottom-up representations learnt via neural networks from emotional facial expressions. Similarly, several recent papers have shown how the linguistic abilities of large language models (LLMs) can be integrated with rational models of planning, communication, and inverse planning (Wong et al., 2023; Ying, Zhi-Xuan, Mansinghka, & Tenenbaum, 2023), modeling human inferences that LLM-based sequence prediction alone struggle with (Zhi-Xuan, Ying, Mansinghka, & Tenenbaum, 2024).

What flexibility does probabilistic programming afford over pure meta-learning? As the article notes, one potential benefit of meta-learning is that it avoids the need for a specific Bayesian model to perform inference over. Crucially, meta-learning achieves this by having access to sufficiently similar data at training and test time, such that the meta-learned algorithm is sufficiently well-adapted to the implied class of data-generating processes. Human cognition is much more adaptive. We do not simply adjust our learning to fit past distributions; we also construct, modify, abstract, and refactor entire *theories* about how the world works (Rule, Tenenbaum, & Piantadosi, 2020; Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Ullman & Tenenbaum, 2020), reasoning with such theories on downstream tasks (Tsividis et al., 2021). This capacity is not captured by pure meta-learning, which occurs "offline." By contrast, probabilistic programming allows modeling these patterns of thought: Theory building can be formulated as program induction (Lake et al., 2015; Saad, Cusumano-Towner, Schaechtle, Rinard, & Mansinghka, 2019), refactoring as program merging (Hwang, Stuhlmüller, & Goodman, 2011), and abstraction-guided reasoning as coarse-to-fine inference (Cusumano-Towner, Bichsel,

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Gehr, Vechev, & Mansinghka, 2018; Stuhlmüller, Hawkins, Siddharth, & Goodman, 2015). Inference meta-programs (Cusumano-Towner et al., 2019; Lew et al., 2023) allow us to model how people invoke modeling and inference strategies as needed: One can employ meta-learned inference when one believes a familiar model applies, but also flexibly compute inferences when a model is learned, extended, or abstracted. On this view, meta-learning has an important role to play in modeling human cognition, but not for all of our cognitive capacities.

Another way of understanding the relationship between meta-learning and probabilistic programming is that the former uses implicit statistical assumptions while the latter's assumptions are explicit. Meta-learning assumes that the structure of the world is conveyed in the statistical structure of data across independent instances. With sufficient coverage of the training distribution, flexible deep learning approaches fit this structure and use it to generalize. But they may not do so in a way that may provide any insight into the *computational* problem being solved by humans. Probabilistic programs, by contrast, explicitly hypothesize the statistical patterns to be found in data, providing constraints that, if satisfied, yield insights for cognition. This implicit–explicit distinction both frames the relative value of the approaches and suggests an alternative relation: A Bayesian model need not subsume or integrate what is learned by a deep learning model, but simply *explicate* it, at a higher level of analysis. Through this lens, having to specify an inference problem is not a limitation, but a virtue.

The best of both worlds will be to compose and further refine these paradigms, such as using deep amortized inference (like meta-learning for Probabilistic Programming), using Bayesian tools (and other tools for mechanistic interpretation) to understand the results of meta-learning, or constructing neurosymbolic models (e.g., by grounding the outputs of meta-learned models in probabilistic programs, as in Wong et al., 2023). As a very recent example, Zhou, Feinman, and Lake (2024) proposed a neurosymbolic program induction model to capture human visual learning, using both Bayesian program induction and meta-learning, achieving the best of both approaches: Interpretability and parsimony, as well as capturing additional variance using flexible function approximators. We believe that the field should move beyond "Connectionist-versus-Bayesian" debates to instead explore hybrid "Connectionist-*and*-Bayesian" approaches.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interest. None.

References

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., & … Goodman, N. D. (2019). Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, *20*(1), 973–978.

Cusumano-Towner, M., Bichsel, B., Gehr, T., Vechev, M., & Mansinghka, V. K. (2018). Incremental inference for probabilistic programs. In Proceedings of the 39th ACM SIGPLAN Conference on Programming Language Design and Implementation (pp. 571–585).

Cusumano-Towner, M. F., Saad, F. A., Lew, A., & Mansinghka, V. K. (2019). Gen: A general-purpose probabilistic programming system with programmable inference. In Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI '19).

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review*, *127*(3), 412.

Goodman, N. D., Mansinghka, V., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2012). Church: a language for generative models. arXiv preprint arXiv:1206.3255.

Goodman N. D., & Stuhlmüller, A. (electronic). The design and implementation of probabilistic programming languages. Retrieved from http://dippl.org.

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.

Hwang, I., Stuhlmüller, A., & Goodman, N. D. (2011). Inducing probabilistic programs by Bayesian program merging. arXiv preprint arXiv:1110.5667.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114.

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

Levy, R., Reali, F., & Griffiths, T. (2008). Modeling the effects of memory on human online sentence processing with particle filters. *Advances in Neural Information Processing Systems*, *21*, 937–944.

Lew, A. K., Matheos, G., Zhi-Xuan, T., Ghavamizadeh, M., Gothoskar, N., Russell, S., & Mansinghka, V. K. (2023). SMCP3: Sequential Monte Carlo with probabilistic program proposals. In International Conference on Artificial Intelligence and Statistics (pp. 7061–7088). PMLR.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348–356.

Ong, D. C., Soh, H., Zaki, J., & Goodman, N. D. (2021). Applying probabilistic programming to affective computing. *IEEE Transactions on Affective Computing*, *12*(2), 306–317.

Rule, J. S., Tenenbaum, J. B., & Piantadosi, S. T. (2020). The child as hacker. *Trends in Cognitive Sciences*, *24*(11), 900–915.

Saad, F. A., Cusumano-Towner, M. F., Schaechtle, U., Rinard, M. C., & Mansinghka, V. K. (2019). Bayesian synthesis of probabilistic programs for automatic data modeling. *Proceedings of the ACM on Programming Languages*, *3*(POPL), 1–32.

Stuhlmüller, A., Hawkins, R. X., Siddharth, N., & Goodman, N. D. (2015). Coarse-to-fine sequential Monte Carlo for probabilistic programs. arXiv preprint arXiv:1509.02962.

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.

Tsividis, P. A., Loula, J., Burga, J., Foss, N., Campero, A., Pouncy, T., & …Tenenbaum, J. B. (2021). Human-level reinforcement learning through theory-based modeling, exploration, and planning. arXiv preprint arXiv:2107.12544.

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*(4), 455–480.

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, *2*, 533–558.

Vul, E., Alvarez, G., Tenenbaum, J., & Black, M. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*, *22*, 1955–1963.

Wong, L., Grand, G., Lew, A. K., Goodman, N. D., Mansinghka, V. K., Andreas, J., & Tenenbaum, J. B. (2023). From word models to world models: Translating from natural language to the probabilistic language of thought. arXiv preprint arXiv:2306.12672.

Ying, L., Zhi-Xuan, T., Mansinghka, V., & Tenenbaum, J. B. (2023). Inferring the goals of communicating agents from actions and instructions. In Proceedings of the AAAI Symposium Series (Vol. 2, No. 1, pp. 26–33).

Zhi-Xuan, T., Ying, L., Mansinghka, V., & Tenenbaum, J. B. (2024). Pragmatic instruction following and goal assistance via cooperative language guided inverse plan search. In Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems.

Zhou, Y., Feinman, R., & Lake, B. M. (2024). Compositional diversity in visual concept learning. *Cognition*, *244*, 105711.

Meta-learning in active inference

O. Penacchio[a]* and A. Clemente[b]

[a]Computer Science Department, Autonomous University of Barcelona, and School of Psychology and Neuroscience, University of St Andrews, Barcelona, Spain and [b]Department of Cognitive Neuropsychology, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
op5@st-andrews.ac.uk
ana.clemente@ae.mpg.de
https://openacchio.github.io/
https://www.aesthetics.mpg.de/institut/mitarbeiterinnen/ana-clemente.html

*Corresponding author.

doi:10.1017/S0140525X24000074, e159

Abstract

Binz et al. propose meta-learning as a promising avenue for modelling human cognition. They provide an in-depth reflection on the advantages of meta-learning over other computational models of cognition, including a sound discussion on how their proposal can accommodate neuroscientific insights. We argue that active inference presents similar computational advantages while offering greater mechanistic explanatory power and biological plausibility.

Binz et al. provide a meritorious survey of the prospects offered by meta-learning for building models of human cognition. Among the main assets of meta-learning discussed by Binz et al. is the capacity to learn inductive bias from experience independently of the constraints enforced by the modeller. Further, Binz et al. showcase the capacity of meta-learning algorithms to approximate Bayesian inference, the gold standard for modelling rational analysis. Finally, they claim that meta-learning offers an unequalled framework for constructing rational models of human cognition that incorporate insights from neuroscience. We propose that an alternative theory of cognition, active inference, shares the same strengths as Binz et al.'s proposal while establishing precise and empirically validated connections to neurobiological mechanisms underlying cognition.

Learning from experience has become a benchmark in all fields aiming to understand and emulate natural intelligence and might be the next driver of developments in artificial intelligence (Zador & Tsao, 2023). In this regard, meta-learning joins other frameworks with the potential to advance the understanding of human cognition as it allows learning algorithms to adapt to experience beyond the modeller's intervention. However, active inference provides a distinct advantage as it has the purpose of modelling and understanding how agents engage with their environment. In active inference, cognitive agents – or algorithms – learn through experience by continually refining their internal model of the environment or the task at hand (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016).

Another critical aspect of Binz et al.'s proposal is their insistent reference to Bayes' optimality. The concept has well-grounded theoretical and empirical foundations (Clark, 2013) that make it a good standard, justifying Binz et al.'s eagerness to probe their approach against Bayes' optimality. Their algorithm approximates Bayes' optimality with the mathematical consequence that any cognitive phenomenon accounted for by Bayesian inference can, in theory, be accounted for by meta-learning. However, the flexibility of Binz et al.'s approach to meta-learning entails a reduced interpretability of the resulting models. In active inference, posterior distributions are inferred using the free-energy principle, a variational approach to Bayesian inference that also approximates intractable computations but within a fully interpretable architecture (see below).

We also commend Binz et al. for describing meta-learning's capacity to incorporate insights from neuroscience, a requisite for a computational understanding of cognition (Kriegeskorte & Douglas, 2018). Yet, the biologically inspired elements introduced are *ad hoc* and case-dependent, as explicitly stated in their conclusion. Their meta-learning models are not motivated by a fundamental biological principle but are conceived as a powerful tool to enhance learning. By contrast, active inference directly translates into neural mechanisms (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017) and originates from a single unifying principle: the imperative for organisms to avoid *surprising* states, implemented by a continuous loop between drawing hypotheses on hidden states (e.g., mean length of an insect species) and observations (e.g., length of a particular specimen) (Friston, 2010). This principle aligns with the Helmoltzian perspective of perception as inference and subsequent Bayesian brain theories. The variational inferential dynamic when receiving new *observations* can be naturally cast into a constant bidirectional *message passing* with direct neural implementation as ascending prediction errors and descending predictions (Pezzulo, Parr, & Friston, 2024). Importantly, these mechanisms are common to all active inference models and enjoy empirical support (e.g., Bastos et al., 2012; Schwartenbeck, FitzGerald, Mathys, Dolan, & Friston, 2015).

Learning is a central construct in active inference. Agents constantly update their generative models based on observations and prediction errors, with the imperative of reducing prediction errors. Generative models represent alternative hypotheses about task execution and associated outcomes (e.g., estimating the average length of a species). These hypotheses, and possibly all components of the generative models, are tested and refined through the agent's experience (Friston et al., 2016). This experience-dependent plasticity has two fundamental assets.

First, learning in active inference is directly and naturally interpreted in terms of biologically plausible neuronal mechanisms. The updates of all components of the generative models are driven by co-occurrences between predicted outcomes (in postsynaptic units in the neuronal interpretation sketched above) and (presynaptic) observational inputs in a process reminiscent of Hebbian learning (Friston et al., 2016). Consequently, active inference is cast as a process theory that can draw specific empirical predictions on neuronal dynamics (Whyte & Smith, 2021).

Second, and crucially, agents in active inference learn the reliability of inputs and prediction errors. This precision estimation is akin to learning meta-parameters (as per Binz et al.) as it entails a weighting process that prioritises reliable sources over uninformative inputs. The balance between exploration and exploitation (central constructs in cognition reflecting epistemic affordances and pragmatic value, respectively) rests upon mechanisms with direct neurobiological substrate in terms of dopamine release, with important implications for rational decision-making – for example, in two-armed bandit tasks (Schwartenbeck et al., 2019), maze navigation (Kaplan & Friston, 2018) and computational psychiatry (Smith, Badcock, & Friston, 2021). Another essential strength of active inference for implementing meta-learning is its natural hierarchical extension. Upper levels can control parameters of lower levels, enabling inference at different timescales whereby learning at lower levels is optimised over time by top-down adjustments from upper levels, which has direct neuronal interpretation in multi-scale hierarchical brain organisation (Pezzulo, Rigoli, & Friston, 2018).

Model preference depends on performance and, primarily, on the scientific question at hand. To understand cognition and its mechanistic underpinnings, models whose components and articulations can be directly interpreted in terms of neural mechanisms are essential. Active inference is a principled, biologically plausible and fully interpretable model of cognition with promising applications to artificial intelligence that accounts for neurobiological and psychological phenomena. We contend that it

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

provides a comprehensive model for understanding biological systems and improving artificial cognition.

Financial support. O. P. was funded by a Maria Zambrano Fellowship for the attraction of international talent for the requalification of the Spanish university system—Next Generation EU.

Competing interests. None.

References

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, *68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, *29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*(4), 323–343. https://doi.org/10.1007/s00422-018-0753-2

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Pezzulo, G., Parr, T., & Friston, K. (2024). Active inference as a theory of sentient behavior. *Biological Psychology*, *186*, 108741. https://doi.org/10.1016/j.biopsycho.2023.108741

Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences*, *22*(4), 294–306. https://doi.org/10.1016/j.tics.2018.01.009

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2015). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral Cortex*, *25*(10), 3434–3445. https://doi.org/10.1093/cercor/bhu159

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H. B., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, *8*, e41703. https://doi.org/10.7554/eLife.41703

Smith, R., Badcock, P., & Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, *75*(1), 3–13. https://doi.org/10.1111/pcn.13138

Whyte, C. J., & Smith, R. (2021). The predictive global neuronal workspace: A formal active inference model of visual consciousness. *Progress in Neurobiology*, *199*, 101918. https://doi.org/10.1016/j.pneurobio.2020.101918

Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., Tsao, D. (2023). Catalyzing next-generation Artificial Intelligence through NeuroAI. *Nature Communications*, *14*(1), 1597. https://doi.org/10.1038/s41467-023-37180-x

Quo vadis, planning?

Jacques Pesnot-Lerousseau[a,b] and Christopher Summerfield[c]*

[a]Institute for Language, Communication, and the Brain, Aix-Marseille Univ, Marseille, France; [b]Aix Marseille Univ, Inserm, INS, Inst Neurosci Syst, Marseille, France and [c]Department of Experimental Psychology, University of Oxford, Oxford, UK
christopher.summerfield@psy.ox.ac.uk
jacques.pesnot-lerousseau@univ-amu.fr
https://humaninformationprocessing.com/

*Corresponding author.

doi:10.1017/S0140525X24000190, e160

Abstract

Deep meta-learning is the driving force behind advances in contemporary AI research, and a promising theory of flexible cognition in natural intelligence. We agree with Binz et al. that many supposedly "model-based" behaviours may be better explained by meta-learning than by classical models. We argue that this invites us to revisit our neural theories of problem solving and goal-directed planning.

The most impressive feats of natural intelligence are the most unfathomable. Caledonian crows fashion hooks to retrieve grubs, honey badgers build ladders to escape from enclosures, and humans have worked out how to split the atom (de Waal, 2016). Humans and other animals are capable of remarkable feats of problem solving in open-ended environments, but we lack computational theories of how this might be achieved (Summerfield, 2022). In the target article, Binz et al. introduce neuroscientists to an exciting new tool: Deep meta-learning. This computational approach provides an interesting candidate solution for some of nature's most startling and puzzling behaviours.

Across the twentieth century, superlative intelligence was synonymous with a capacity for planning, so early AI researchers believed that if a machine ever vanquished a human at chess, then AI would have been solved. Classical models conceive of the world as a list of states and their transition probabilities; planning requires efficient *search*, or mental exploration of possible pathways to reach a goal. Neuroscientists still lean heavily on these classical models to understand how rodents and primates solve problems like navigating to a spatial destination, assuming that these rely on "model-based" search or forms of offline rumination (Daw & Dayan, 2014). However, in contemporary machine learning, new explanations of complex sequential behaviours are emerging.

Today – nearly three decades since the first electronic chess grandmaster – AI research is dominated by deep network models, which exploit massive training datasets to learn complex functions mapping inputs onto outputs. In fact, in machine learning, explicitly model-based solutions to open-ended problems have not lived up to their promise. To give one example: In 1997, the chess program Deep Blue defeated the world champion Garry Kasparov using tree search with an alpha–beta search algorithm, ushering in an era where computers played stronger chess than people. In 2017, DeepMind's hybrid network AlphaZero defeated chess computer champion Stockfish by augmenting the search algorithm (Monte Carlo Tree Search) with a deep neural network that learned from (self-)play to evaluate board positions (Silver et al., 2018). In early 2024, performance comparable to the best human players was achieved using a deep network alone without search, thanks to computational innovation (transformer networks) and increasing scale (millions of parameters). AI research has implied that when big brains are exposed to big data, explicit forms of lookahead play a limited role in their success. The authors encapsulate this view with a quote attributed to 1920s grandmaster José Raúl Capablanca: "*I see only one move ahead, but it is always the correct one*" (Ruoss et al., 2024).

Deep meta-learning applies powerful function approximation to sequential decision problems, where optimal policies may involve forms of exploration or active hypothesis testing to meet long-term objectives. In the domain of reinforcement learning (RL), deep meta-RL can account for human behaviours on

benchmark problems thought to tap into model-based inference or planning, such as the "two-step" decision task, without invoking the need for search. This is because a deep neural network equipped with a stateful activation memory, and meta-trained to a wide range of sequential decision problems, can learn a policy that is intrinsically cognitively flexible. It learns to react on the fly to the twists and turns of novel sequential environments, and thus produce the sorts of behaviours that were previously thought to be possible only with model-based forms of inference. Paradoxically, although "meta-learning" means "learning to learn," inner loop learning can occur in "frozen" networks – those without parameter updates. This offers a plausible model of how recurrent neural systems for memory and control, housed in prefrontal cortex, allow us to solve problems we have never seen before without explicit forms of search (Wang et al., 2018).

In AI research today, the most successful deep meta-learning systems are large transformer-based networks that are trained to complete sequences of tokenised natural language (Large Language Models or LLMs). These networks learn semantic and syntactic patterns that allow them to solve a very open-ended problem – constructing a relevant, coherent sentence. Researchers working with LLMs today call deep meta-learning "in context learning" because instead of using recurrent memory, transformers are purely feedforward networks that rely on autoregression – past outputs are fed back in as inputs, providing a context on which to condition the generative process. Although transformers do not resemble plausible neural algorithms, their striking success has opened up new questions concerning neural computation. For example, in-context learning proceeds faster when exemplar ordering is structured rather than random, like human learning but unlike traditional in-weight learning (Chan et al., 2022b), and "in-context learning" may be better suited to rule learning than in-weight learning (Chan et al., 2022a). Meta-learning thus offers new tools for psychologists and neuroscientists interested in biological learning and memory in natural agents.

Binz et al. argue that we should take deep meta-learning seriously as an alternative to Bayesian decision theory. We would go further, arguing that meta-learning is a candidate general theory for flexible cognition. It explains why executive function improves dramatically with experience (Ericsson & Charness, 1994). Unlike Deep Blue, human world chess no.1 Magnus Carlsen has improved since his first game at the age of five. Purely search-based accounts flexible cognition are obliged to propose that performance will plateau as soon as the transition function (e.g., the rules of chess) are fully mastered, or else posit unexplained ways in which search policies deepen or otherwise mutate with practice (Van Opheusden et al., 2023). In a world where states are heterogeneous and noisy, deep meta-learning explains how we can generalise sequential behaviours to novel states. In a world where speed and processing power are at a premium, meta-learning shifts the burden of inference to the training period, allowing for fast and efficient online computation. Meta-learning is a general theory of natural intelligence that is – more than classical counterpart – fit for the real world.

Undoubtedly, humans and other animals do engage in explicit forms of planning, especially when the stakes are high. But many sequential behaviours that are thought to index this ability may rely more on deep meta-learning than classical planning.

Financial support. This work was supported by Fondation Pour l'Audition FPA RD-2021-2 (J. P.-L.), Institute for Language, Communication, and the Brain ILCB (J. P.-L.) and Wellcome Trust Discovery Award (227928/Z/23/Z) to (C. S.).

Competing interests. None.

References

Chan, S. C. Y., Dasgupta, I., Kim, J., Kumaran, D., Lampinen, A. K., & Hill, F. (2022a). Transformers generalize differently from information stored in context vs in weights. https://doi.org/10.48550/ARXIV.2210.05675

Chan, S. C. Y., Santoro, A., Lampinen, A. K., Wang, J. X., Singh, A., Richemond, P. H., … Hill, F. (2022b). Data distributional properties drive emergent in-context learning in transformers. https://doi.org/10.48550/ARXIV.2205.05055

Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B*, *369*, 20130478. https://doi.org/10.1098/rstb.2013.0478

de Waal, F. (2016). *Are we smart enough to know how smart animals are?* W. W. Norton & Company.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*, 725–747. https://doi.org/10.1037/0003-066X.49.8.725

Ruoss, A., Delétang, G., Medapati, S., Grau-Moya, J., Wenliang, L. K., Catt, E., … Genewein, T. (2024). Grandmaster-level chess without search.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., … Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*, 1140–1144. https://doi.org/10.1126/science.aar6404

Summerfield, C. (2022). *Natural general intelligence: How understanding the brain can help us build AI*. Oxford University Press.

Van Opheusden, B., Kuperwajs, I., Galbiati, G., Bnaya, Z., Li, Y., & Ma, W. J. (2023). Expertise increases planning depth in human gameplay. *Nature*, *618*, 1000–1005. https://doi.org/10.1038/s41586-023-06124-2

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J.Z., … Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, *21*, 860–868. https://doi.org/10.1038/s41593-018-0147-8

The hard problem of meta-learning is what-to-learn

Yosef Prat* and Ehud Lamm

The Cohn Institute for History and Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv, Israel
yosefprat@gmail.com
ehudlamm@post.tau.ac.il
https://www.ehudlamm.com

*Corresponding author.

doi:10.1017/S0140525X24000268, e161

Abstract

Binz et al. highlight the potential of meta-learning to greatly enhance the flexibility of AI algorithms, as well as to approximate human behavior more accurately than traditional learning methods. We wish to emphasize a basic problem that lies underneath these two objectives, and in turn suggest another perspective of the required notion of "meta" in meta-learning: knowing what to learn.

We postulate that the hard problem in (natural or artificial) intelligence is the question of "what to learn?". At the fundamental level, this meta-question is resolved in nature by the evolutionary process. The question of "how to learn?", which is the focus of the meta-learning framework that is presented in the target article, is not especially easy as well, but it can be captured by devising

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

specific training structures and relevant optimization tasks. In general, it requires to specify the "search space" of possible learning strategies. The hard problem of learning, however, is the identification of the learning task itself (i.e., what, and if, to learn) (Niv, 2019). For instance, a real-life learner observing several specimens of some unknown insect species (following the example in the target article) must first somehow realize that she is required to evaluate the average length of that species, before she begins to tune her evaluation strategies. This is indeed a different meta-task than presented by Binz et al., but its solution is mandatory for any artificial (somewhat-)general intelligence, and it is regularly handled by the brain (Roli, Jaeger, & Kauffman, 2022).

In the quest to devise domain-general learning models, Binz et al. correctly identify the need for diverse (and maybe realistic) training sets. Training a model on many different tasks can achieve high performance in all of them, and maybe even in unrelated, but similar, tasks. Yet, the model will always be constrained by the task-space spanned by its training sets. The major challenge does not lie in amplifying the dimensionality, or variability, of the learned problem, but rather in determining the appropriate objective function. Here, we may be inspired by the observation that biological brains have in general not evolved for their ability to solve a specific task, but, rather, are shaped by the overall success of the organism. On the one hand, evolutionary success obscures the objective of each specific task, since it depends on long-term benefits that are not always clearly related to short-term behavior. On the other hand, evolutionary success is a broader optimization challenge. A generic model that can both solve a maze and evaluate the average length of a newly identified insect, without being trained specifically on these tasks, must solve the hard problem of what-to-learn in a given context. To build a model that addresses this challenge we cannot handcraft the utility function (or error measurements) of each task separately. The meta-learning requirement thus becomes to learn how to identify the utility in learning, or in performing, each of the given tasks, and more broadly, to identify the task itself. Thus, it is constraining to use training sets and error functions that provide the learner with "correct" answers or feedback for each task separately, as is typically done in supervised, semi-supervised, and reinforcement machine learning. The biological brain is overall domain-general since it is not guided by a task-specific "utility function." Domain-specific processes, such as, maybe, those suggested to process language (Fedorenko & Blank, 2020), demonstrate cases in which natural selection narrowed or "optimized" the task of finding what-to-learn. Other indications may include modularity (Ellefsen, Mouret, & Clune, 2015; Sporns & Betzel, 2016), alongside sensory adaptations (Warrant, 2016), attention biases (Niv et al., 2015), and data acquisition mechanisms (Lotem & Halpern, 2012). Furthermore, in humans, cultural evolution may also adjust task specificity (Heyes, 2018).

The evolutionary process may also explain the limitations of treating cognition as rational, or optimal. Binz et al. suggest that unrealistic aspects of Bayesian models can be mitigated using resource constraints, for which the offered meta-learning framework is suitable. The problem, however, is that human (and other animal) behavior is not straightforwardly rational, and often appears to defy Bayesian optimization (Tversky & Kahneman, 1981). Moreover, this may not be due to limited resources but because the success of living creatures is determined evolutionarily, rather than by immediate outcomes (Houston, McNamara, & Steer, 2007). When behavioral objectives are considered on an evolutionary scale, it may be revealed that they are (locally) optimal (Kacelnik, 2006), and this includes behaviors that depend upon learning, as is generally assumed in behavioral ecology. When tasks for which learning is evolutionarily beneficial end up being learned (i.e., when those individuals who learn have higher fitness), natural selection resolves the meta-learning hard problem of what-to-learn (Dunlap & Stephens, 2016). This may bias the things that animals are able to learn, by shaping the parameter search-space (Prat, Bshary, & Lotem, 2022), maybe of the outer learning loop described by Binz et al. These biases are often addressed in the biological learning literature as sub-problems of the what-to-learn problem, and include when-to-learn or from whom-to-learn (Laland, 2004).

We suggest that further advancements in meta-learning thinking require addressing the hard problem of learning as one of their aims. Inspired by (human and nonhuman) biological brains, this should be done by devising overarching objectives for learning algorithms that will enable them to learn what are the learning tasks. In nature, evolution provides some of the solution. Yet, it is not necessary to mimic the evolutionary process per se, but only to acknowledge the generality of evolutionary optimization in the natural world. To this end, it may be better to aspire to simulate nonhuman-animal behavioral studies, rather than psychological assays, since nonhuman animals are trained with no description of the boundaries of their task – they need to realize it by themselves (e.g., when a sparrow learns to relate sand color to food [Ben-Oren, Truskanov, & Lotem, 2022]). Thus, these studies usually contain a direct meta-learning challenge that requires solving the problem of what-to-learn.

Acknowledgements. We thank Yoav Ram for insightful comments on a previous version of the manuscript.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References

Ben-Oren, Y., Truskanov, N., & Lotem, A. (2022). House sparrows use learned information selectively based on whether reward is hidden or visible. *Animal Cognition*, *25*(6), 1545–1555. https://doi.org/10.1007/s10071-022-01637-1

Dunlap, A. S., & Stephens, D. W. (2016). Reliability, uncertainty, and costs in the evolution of animal learning. *Current Opinion in Behavioral Sciences*, *12*, 73–79. https://doi.org/https://doi.org/10.1016/j.cobeha.2016.09.010

Ellefsen, K. O., Mouret, J.-B., & Clune, J. (2015). Neural modularity helps organisms evolve to learn new skills without forgetting old skills. *PLoS Computational Biology*, *11*(4), e1004128. https://doi.org/doi.org/10.1371/journal.pcbi.1004128

Fedorenko, E., & Blank, I. A. (2020). Broca's area is not a natural kind. *Trends in Cognitive Sciences*, *24*(4), 270–284. https://doi.org/10.1016/j.tics.2020.01.001

Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press. https://doi.org/10.2307/j.ctv24trbqx

Houston, A. I., McNamara, J. M., & Steer, M. D. (2007). Do we expect natural selection to produce rational behaviour? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1531–1543. https://doi.org/10.1098/rstb.2007.2051

Kacelnik, A. (2006). Meanings of rationality. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 87–106). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528272.003.0002

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14. https://doi.org/10.3758/BF03196002

Lotem, A., & Halpern, J. Y. (2012). Coevolution of learning and data-acquisition mechanisms: A model for cognitive evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2686–2694. https://doi.org/10.1098/rstb.2012.0213

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on

attention mechanisms. *The Journal of Neuroscience*, *35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Prat, Y., Bshary, R., & Lotem, A. (2022). Modelling how cleaner fish approach an ephemeral reward task demonstrates a role for ecologically tuned chunking in the evolution of advanced cognition. *PLoS Biology*, *20*(1), e3001519. https://doi.org/10.1371/journal.pbio.3001519

Roli, A., Jaeger, J., & Kauffman, S. A. (2022). How organisms come to know the world: Fundamental limits on artificial general intelligence. *Frontiers in Ecology and Evolution*, *9*, 806283. https://doi.org/10.3389/fevo.2021.806283

Sporns, O., & Betzel, R. F. (2016). Modular brain networks. *Annual Review of Psychology*, *67*(1), 613–640. https://doi.org/10.1146/annurev-psych-122414-033634

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Warrant, E. J. (2016). Sensory matched filters. *Current Biology*, *26*(20), R976–R980. https://doi.org/https://doi.org/10.1016/j.cub.2016.05.042

Is human compositionality meta-learned?

Jacob Russin[a,b], Sam Whitman McGrath[c], Ellie Pavlick[a] and Michael J. Frank[d]*

[a]Department of Computer Science, Brown University, Providence, RI, USA; [b]Department of Cognitive and Psychological Sciences, Brown University, Providence, RI, USA; [c]Department of Philosophy, Brown University, Providence, RI, USA and [d]Department of Cognitive and Psychological Sciences, Carney Institute for Brain Science, Brown University, Providence, RI, USA

jake_russin@brown.edu
sam_mcgrath1@brown.edu
ellie_pavlick@brown.edu
michael_frank@brown.edu
https://jlrussin.github.io/
https://scholar.google.com/citations?user=B3b7kAYAAAAJ&hl=en
https://cs.brown.edu/people/epavlick/
http://ski.clps.brown.edu/

*Corresponding author.

doi:10.1017/S0140525X24000189, e162

Abstract

Recent studies suggest that meta-learning may provide an original solution to an enduring puzzle about whether neural networks can explain compositionality – in particular, by raising the prospect that compositionality can be understood as an emergent property of an inner-loop learning algorithm. We elaborate on this hypothesis and consider its empirical predictions regarding the neural mechanisms and development of human compositionality.

Binz et al. review recent meta-learned models that can reproduce human-like compositional generalization behaviors (Lake & Baroni, 2023), but they stop short of endorsing meta-learning as a theoretical framework for understanding human compositionality. Here, we elaborate on this proposal, articulating the hypothesis that human compositionality can be understood as an emergent property of an inner-loop, in-context learning algorithm that is itself meta-learned.

Compositionality has played a key theoretical role in cognitive science since its inception (Chomsky, 1957), providing an explanation for human systematic and productive generalization behaviors. These phenomena are readily explained by the compositionality of classical cognitive architectures, as the design of their symbolic representations and structure-sensitive operations intrinsically guarantees that they can redeploy familiar constituents in novel constructions (Fodor & Pylyshyn, 1988). It has been argued that neural networks are in principle incapable of playing the same explanatory role because they lack these architectural features (Fodor & Pylyshyn, 1988; Marcus, 1998).

Much work has explored inductive biases that might encourage compositionality to emerge in neural networks (Russin, Jo, O'Reilly, & Bengio, 2020a; Smolensky, 1990; Webb et al., 2024), but meta-learning offers an original solution to the puzzle. As Binz et al. emphasize, when an inner-loop, in-context learning algorithm emerges within the activation dynamics of a meta-learning neural network, it can have fundamentally different properties than the outer-loop algorithm. Thus, even if the outer-loop algorithm lacks these inductive biases, the network may nevertheless implement an emergent in-context learning algorithm that embodies them implicitly.

Lake and Baroni (2023) have shown that such an inner-loop algorithm can pass tests of compositionality that standard neural networks fail (Lake & Baroni, 2018). The question, then, is whether such networks can serve as explanatory models of human compositional generalization. Can we think of human compositionality as an emergent property of an inner-loop, in-context learning algorithm? How might we evaluate such a hypothesis? Here, we consider two independent aspects of this proposal: First, its implications for neural mechanisms, and second, for development.

One straightforward mechanistic prediction is that employing inner-loop, in-context learning mechanisms, rather than outer-loop learning mechanisms, should facilitate compositional generalization behaviors. Cognitive and computational neuroscience provides empirical support for this prediction. Cognitive control – the ability to overcome existing prepotent responses and to flexibly adapt to arbitrary goals (Miller & Cohen, 2001) – is an important capacity for human in-context learning. The neural mechanisms known to be involved in cognitive control, such as working memory, gating, and top-down modulation in the prefrontal cortex (Miller & Cohen, 2001; O'Reilly & Frank, 2006; Russin, O'Reilly, & Bengio, 2020b), are also thought to be essential to compositional abilities such as inferring and applying rules (Calderon, Verguts, & Frank, 2022; Collins & Frank, 2013; Frank & Badre, 2012; Kriete, Noelle, Cohen, & O'Reilly, 2013), deductive and inductive reasoning (Crescentini et al., 2011; Goel, 2007), and processing complex syntax (Thompson-Schill, 2005). Thus, a shared set of neural mechanisms may underlie both in-context learning and compositionality in humans, lending support to the meta-learning hypothesis.

A second, independent prediction is a developmental one – that human compositional generalization abilities are themselves meta-learned over the course of development. Adults come into any psychological experiment equipped with a wealth of prior experience. The meta-learning hypothesis predicts that this includes experiences encouraging the adoption of more compositional learning strategies (i.e., ones sensitive to implicit compositional structure). In general, children exhibit a developmental trajectory consistent with this hypothesis. Older children learn new tasks more efficiently (Bergelson, 2020), especially when these tasks involve cognitive capacities essential to in-context learning, such as working memory and executive functions (Munakata, Snyder, & Chatham, 2012). Furthermore, children

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

improve throughout development on tasks involving the composition of rules (Piantadosi & Aslin, 2016; Piantadosi, Palmeri, & Aslin, 2018).

Innate mechanisms or inductive biases may still be required to successfully meta-learn a compositional inner-loop algorithm in the first place. Indeed, studies in machine learning have shown that architecture seems to be an important factor in determining whether in-context learning capabilities emerge (Chan et al., 2022). Similarly, findings from cognitive and computational neuroscience have emphasized the importance of architectural features such as prefrontal gating mechanisms for the emergence of abstract representations that could mediate subsequent in-context generalization abilities (Collins & Frank, 2013; Frank & Badre, 2012; Kriete et al., 2013; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005). These inductive biases can also explain incidental hierarchical rule learning and generalization in infants (Werchan, Collins, Frank, & Amso, 2015, 2016). Thus, a combination of innate architectural features and meta-learning experiences may be necessary for human compositionality to emerge.

The meta-learning datasets used in previous modeling efforts have typically been developmentally unrealistic because they have been contrived to engender narrow compositional generalization abilities that are specific to a particular type of task. Could meta-learning in less explicitly structured learning scenarios lead to the acquisition of broader compositional generalization abilities? This question deserves careful empirical study, but we may draw a preliminary insight from the success of large language models (Brown et al., 2020), which develop in-context learning abilities (von Oswald et al., 2023; Xie, Raghunathan, Liang, & Ma, 2022) that in some cases exhibit human-like compositionality (Webb, Holyoak, & Lu, 2022; Wei et al., 2023; Zhou et al., 2022). Unlike models explicitly designed for meta-learning, large language models are trained to predict the next token on very large datasets of unstructured text. These datasets contain more language data than humans are exposed to in an entire lifetime (Linzen & Baroni, 2021), so future work needs to investigate what kinds of inductive biases are necessary to improve their sample efficiency. However, these models provide proof of concept that neural networks can develop compositional in-context learning algorithms by training on relatively unstructured data.

Binz et al. shy away from a robust commitment to meta-learning as a theoretical framework, instead emphasizing its utility as a methodological tool. Here, we have demonstrated how the meta-learning perspective on human compositionality can generate testable empirical hypotheses about underlying mechanisms and developmental trajectory. If such a research program bears fruit, it will elevate meta-learning from a useful tool to a novel cognitive theory.

Financial support. M. J. F. is supported by ONR grant N00014-23-1-2792. E. P. and J. R. are supported by COBRE grant no. 5P20GM103645-10.

Competing interest. None.

References

Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child Development Perspectives*, *14*(3), 142–149. https://doi.org/10.1111/cdep.12373

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P.. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Calderon, C. B., Verguts, T., & Frank, M. J. (2022). Thunderstruck: The ACDC model of flexible sequences and rhythms in recurrent neural circuits. *PLoS Computational Biology*, *18*(2), e1009854. https://doi.org/10.1371/journal.pcbi.1009854

Chan, S. C. Y., Santoro, A., Lampinen, A. K., Wang, J. X., Singh, A., Richemond, P. H., … Hill, F. (2022). Data distributional properties drive emergent in-context learning in transformers. *Advances in Neural Information Processing Systems*, *35*, 18878–18891. https://papers.nips.cc/paper_files/paper/2022/hash/77c6ccacfd9962e2307fc64680fc5ace-Abstract-Conference.html

Chomsky, N. (Ed.). (1957). *Syntactic structures*. Mouton & Co.

Collins, A. G. E., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. https://doi.org/10.1037/a0030852

Crescentini, C., Seyed-Allaei, S., De Pisapia, N., Jovicich, J., Amati, D., & Shallice, T. (2011). Mechanisms of rule acquisition and rule following in inductive reasoning. *Journal of Neuroscience*, *31*(21), 7763–7774. https://doi.org/10.1523/JNEUROSCI.4579-10.2011

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1–2), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex*, *22*(3), 509–526. https://doi.org/10.1093/cercor/bhr114

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences*, *11*(10), 435–441. https://doi.org/10.1016/j.tics.2007.09.003

Kriete, T., Noelle, D. C., Cohen, J. D., & O'Reilly, R. C. (2013). Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16390–16395. https://doi.org/10.1073/pnas.1303547110

Lake, B. M., & Baroni, M. (2018). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In J. G. Dy & A. Krause (Eds.), *Proceedings of the 35th International Conference on Machine Learning* (Vol. 80, pp. 2879–2888). PMLR. http://proceedings.mlr.press/v80/lake18a.html

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*, 1–7. https://doi.org/10.1038/s41586-023-06668-3

Linzen, T., & Baroni, M. (2021). Syntactic structure from deep learning. *Annual Review of Linguistics*, *7*(1), 195–212. https://doi.org/10.1146/annurev-linguistics-032020-051035

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, *37*(3), 243–282. https://doi.org/10.1006/cogp.1998.0694

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: Three key transitions. *Current Directions in Psychological Science*, *21*(2), 71–77. https://doi.org/10.1177/0963721412436807

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

Piantadosi, S., & Aslin, R. (2016). Compositional reasoning in early childhood. *PLoS ONE*, *11*(9), e0147734. https://doi.org/10.1371/journal.pone.0147734

Piantadosi, S. T., Palmeri, H., & Aslin, R. (2018). Limits on composition of conceptual operations in 9-month-olds. *Infancy*, *23*(3), 310–324. https://doi.org/10.1111/infa.12225

Rougier, N. P., Noelle, D., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2005). Prefrontal cortex and the flexibility of cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7338–7343.

Russin, J., Jo, J., O'Reilly, R. C., & Bengio, Y. (2020a). Systematicity in a recurrent neural network by factorizing syntax and semantics. *Proceedings for the 42nd Annual Meeting of the Cognitive Science Society*, 7. https://cognitivesciencesociety.org/cogsci20/papers/0027/0027.pdf

Russin, J., O'Reilly, R. C., & Bengio, Y. (2020b). Deep learning needs a prefrontal cortex. In Bridging AI and Cognitive Science (BAICS) Workshop, ICLR, 2020, 11.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*(1–2), 159–216. https://doi.org/10.1016/0004-3702(90)90007-M

Thompson-Schill, S. L. (2005). Dissecting the language organ: A new look at the role of Broca's area in language processing. In Anne Cutler (Ed.), *Twenty-first century psycholinguistics* (1st ed., Vol. 1, pp. 1–18). Routledge.

von Oswald, J., Niklasson, E., Schlegel, M., Kobayashi, S., Zucchet, N., Scherrer, N., … Sacramento, J. (2023). Uncovering mesa-optimization algorithms in transformers (arXiv:2309.05858). arXiv. https://doi.org/10.48550/arXiv.2309.05858

Webb, T., Frankland, S. M., Altabaa, A., Krishnamurthy, K., Campbell, D., Russin, J., … Cohen, J. D. (2024). The relational bottleneck as an inductive bias for efficient abstraction (arXiv:2309.06629). arXiv. http://arxiv.org/abs/2309.06629

Webb, T., Holyoak, K. J., & Lu, H. (2022). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, *7*(9). https://doi.org/10.1038/s41562-023-01659-w

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., … Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837. https://papers.nips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf

Werchan, D. M., Collins, A. G. E., Frank, M. J., & Amso, D. (2015). 8-Month-old infants spontaneously learn and generalize hierarchical rules. *Psychological Science*, *26*(6), 805–815. https://doi.org/10.1177/0956797615571442

Werchan, D. M., Collins, A. G. E., Frank, M. J., & Amso, D. (2016). Role of prefrontal cortex in learning and generalizing hierarchical rules in 8-month-old infants. *The Journal of Neuroscience*, *36*(40), 10314–10322. https://doi.org/10.1523/JNEUROSCI.1351-16.2016

Xie, S. M., Raghunathan, A., Liang, P., & Ma, T. (2022). An explanation of in-context learning as implicit Bayesian inference. *International Conference on Learning Representations*. https://openreview.net/pdf?id=RdJVFCHjUMI

Zhou, D., Schärli, N., Hou, L., Wei, J., Scales, N., Wang, X., … Chi, E. (2022). Least-to-most prompting enables complex reasoning in large language models. *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=WZH7099tgfM

Combining meta-learned models with process models of cognition

Adam N. Sanborn*, Haijiang Yan and Christian Tsvetkov

Department of Psychology, University of Warwick, Coventry, UK
a.n.sanborn@warwick.ac.uk
haijiang.yan@warwick.ac.uk
chris.tsvetkov@warwick.ac.uk
https://go.warwick.ac.uk/adamsanborn

*Corresponding author.

doi:10.1017/S0140525X24000165, e163

Abstract

Meta-learned models of cognition make optimal predictions for the actual stimuli presented to participants, but investigating judgment biases by constraining neural networks will be unwieldy. We suggest combining them with cognitive process models, which are more intuitive and explain biases. Rational process models, those that can sequentially sample from the posterior distributions produced by meta-learned models, seem a natural fit.

Meta-learned models of cognition offer an exciting opportunity to address a central weakness of current cognitive models, whether Bayesian or not: Cognitive models generally do not "see" the experimental stimuli shown to participants. Experimenters instead feed models low-dimensional descriptions of the stimuli, which are often in terms of the psychological features imagined by the experimenter, or sometimes are the psychological descriptions that best fit participants' judgments (e.g., stimulus similarity judgments; Nosofsky, Sanders, Meagher, & Douglas, 2018).

For example, in studies of probability judgment, participants have been asked to judge the probability that "Bill plays jazz for a hobby" after having been given the description, "Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities" (Tversky & Kahneman, 1983). Current probability judgment models reduce these descriptions down to a single unknown number, and attempt to find the latent probability that best fits the data (e.g., Zhu, Sanborn, & Chater, 2020).

Models trained on the underlying statistics of the environment, as meta-learned models are, can bypass this need to infer a latent variable, instead making predictions from the actual descriptions used. Indeed, even relatively simple models of semantics that locate phrases in a vector space produce judgments that correlate with the probabilities experimental participants give (Bhatia, 2017). Meta-learned models could thus explain a great deal of the variability in human behavior, and allow experimenters to generalize beyond the stimuli shown to participants.

However, used as descriptive models, normative meta-learned models of cognition inherit a fundamental problem from the Bayesian approach: People's reliable deviations from normative behavior. One compelling line of research shows that probability judgments are incoherent in a way that Bayesian models are not. Using the above example of Bill, Tversky and Kahneman (1983) found participants ranked the probability of "Bill is an accountant who plays jazz for a hobby" as higher than that of "Bill plays jazz for a hobby." This violates the extension rule of probability because the set of all accountants who play jazz for a hobby is a subset of all people who play jazz for a hobby, no matter how Bill is described.

The target article discusses constraining meta-learned models to better describe behavior, such as reducing the number of hidden units or restricting the representational fidelity of units. These manipulations have produced a surprising and interesting range of biases, including stochastic and incoherent probability judgments (Dasgupta, Schulz, Tenenbaum, & Gershman, 2020). However, this is just the start to explaining human biases. Even a single bias such as the conjunction fallacy has intricacies, such as the higher rate of conjunction fallacies when choosing versus estimating (Wedell & Moro, 2008), and greater variability in judgments of conjunctions than those of simple events (Costello & Watts, 2017).

Cognitive process models aim to explain these biases in detail. For conjunction fallacies, a variety of well-supported models exist, based on ideas such as participants sampling events with noise in the retrieval process (Costello & Watts, 2014), or by sacrificing probabilistic coherence to improve judgment accuracy based on samples (Zhu et al., 2020), or by representing conjunctions as a weighted average of simple events (Juslin, Nilsson, & Winman, 2009), or by using quantum probability (Busemeyer, Pothos, Franco, & Trueblood, 2011). These kinds of models capture many details of the empirical effects, through simple and intuitive mechanisms like adjusting the amount of noise or number of samples, which helps identify experiments to distinguish between them.

Mechanistically modifying meta-learned models to explain cognitive biases to the level cognitive process models do appears difficult. While changes to network structure are powerful ways to induce different biases that could identify implementation-level constraints in the brain, the effects of these kinds of changes are generally hard to intuit, while training constrained meta-learning models to test different manipulations will be slow and computationally expensive. Thus, it will be challenging to reproduce existing biases in detail or to design effective experiments for testing these constraints.

Combining meta-learned models with cognitive process models is more promising. One possibility is to have meta-learned models act as a "front end" that takes stimuli and converts them to a feature-based representation, which is then operated on by a cognitive process model. The parameters of the cognitive process model could be fit to human data, or potentially the cognitive process model could be encoded into the network (e.g.,

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Peterson, Bourgin, Agrawal, Reichman, & Griffiths, 2021), and meta-learning could be done on the front end and the cognitive process parameters end-to-end.

However, as meta-learned models of cognition produce posterior predictive distributions, rational process models offer a straightforward connection that does not require retraining meta-learned models. Rational process models do not directly use a posterior predictive distribution, but instead assume that the posterior predictive distribution is approximated (i.e., using the posterior mean, posterior median, or other summary statistic depending on task), most often using a statistical sampling algorithm (Griffiths, Vul, & Sanborn, 2012). Such a model can explain details of the conjunction fallacy, and also a wide range of other biases, such as stochastic choice, anchoring and repulsion effects in estimates, long-range autocorrelations in judgment, and the flaws in random sequence generation (Castillo, León-Villagrá, Chater, & Sanborn, 2024; Spicer, Zhu, Chater, & Sanborn, 2022; Vul, Goodman, Griffiths, & Tenenbaum, 2014; Zhu, León-Villagrá, Chater, & Sanborn, 2022, 2023). What these models have lacked, however, is a principled way to construct the posterior predictive distribution from environmental statistics, and here meta-learned models offer that exciting possibility.

While rational process models offer what we think is a natural choice for integration, any sort of combination with existing cognitive models offers benefits. Being able to explain both the details of biases as cognitive process models do, as well as showing sensitivity to actual stimuli is a powerful combination that moves toward the long-standing goal of a general model of cognition. Overall we see meta-learned models of cognition as not supplanting existing cognitive models, but as a way to make them much more powerful and relevant to understanding and predicting behavior.

Acknowledgments. None.

Financial support. A. N. S. and C. T. were supported by a European Research Council consolidator grant (817492 – SAMPLING). H. Y. was supported by a Chancellor's International Scholarship from the University of Warwick.

Competing interest. None.

References

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20. http://dx.doi.org/10.1037/rev0000047

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review*, *118*(2), 193–218. https://doi.org/10.1037/a0022542

Castillo, L., León-Villagrá, P., Chater, N., & Sanborn, A. (2024). Explaining the flaws in human random generation as local sampling with momentum. *PLoS Computational Biology*, *20*(1), e1011739. https://doi.org/10.1371/journal.pcbi.1011739

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*(3), 463–480. https://doi.org/10.1037/a0037010

Costello, F., & Watts, P. (2017). Explaining high conjunction fallacy rates: The probability theory plus noise account. *Journal of Behavioral Decision Making*, *30*(2), 304–321. https://dx.doi.org/10.1002/bdm.1936

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review*, *127*(3), 412–441. https://doi.org/10.1037/rev0000178

Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, *21*(4), 263–268. https://doi.org/10.1177/0963721412447619

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review*, *116*(4), 856–874. https://doi.org/10.1037/a0016979

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior Research Methods*, *50*, 530–556. https://doi.org/10.3758/s13428-017-0884-8

Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D., & Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, *372*(6547), 1209–1214. https://doi.org/10.1126/science.abe2629

Spicer, J., Zhu, J. Q., Chater, N., & Sanborn, A. N. (2022). Perceptual and cognitive judgments show both anchoring and repulsion. *Psychological Science*, *33*(9), 1395–1407. https://doi.org/10.1177/09567976221089599

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. https://doi.org/10.1037/0033-295X.90.4.293

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637. https://doi.org/10.1111/cogs.12101

Wedell, D. H., & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus, and problem type. *Cognition*, *107*(1), 105–136. https://doi.org/10.1016/j.cognition.2007.08.003

Zhu, J. Q., León-Villagrá, P., Chater, N., & Sanborn, A. N. (2022). Understanding the structure of cognitive noise. *PLoS Computational Biology*, *18*(8), e1010312. https://doi.org/10.1371/journal.pcbi.1010312

Zhu, J.-Q., Sanborn, A. N., & Chater, N. (2020). The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review*, *127*(5), 719–748. https://doi.org/10.1037/rev0000190

Zhu, J.-Q., Sundh, J., Spicer, J., Chater, N., & Sanborn, A. N. (2023). The autocorrelated Bayesian sampler: A rational process for probability judgments, estimates, confidence intervals, choices, confidence judgments, and response times. *Psychological Review*, *131*(2), 456–493. https://doi.org/10.1037/rev0000427

Linking meta-learning to meta-structure

Malte Schilling[a]*, Helge J. Ritter[b] and Frank W. Ohl[c,d]

[a]Autonomous Intelligent Systems Group, Computer Science Department, University of Münster, Münster, Germany; [b]Neuroinformatics Group Faculty of Technology/CITEC, Bielefeld University, Bielefeld, Germany; [c]Department of Systems Physiology of Learning, Leibniz Institute for Neurobiology, Magdeburg, Germany and [d]Institute of Biology, Otto-von-Guericke University, Magdeburg, Germany

malte.schilling@uni-muenster.de
helge@techfak.uni-bielefeld.de
frank.ohl@lin-magdeburg.de
https://www.uni-muenster.de/AISystems/
https://ni.www.techfak.uni-bielefeld.de/people/helge/
https://www.ovgu.de/Ohl.html

*Corresponding author.

doi:10.1017/S0140525X24000232, e164

Abstract

We propose that a principled understanding of meta-learning, as aimed for by the authors, benefits from linking the focus on learning with an equally strong focus on structure, which means to address the question: What are the meta-structures that can guide meta-learning?

The authors discuss meta-learning as a flexible and computationally efficient tool to generate cognitive models from training data and thereby to avoid the need for handcrafting cognitive biases as usually done in current cognitive architectures or Bayesian learning. They provide four supporting arguments as a motivation for a systematic research program on meta-learning, which they diagnose as so far largely missing. While we agree with this stance, we propose that a deeper understanding of meta-learning would benefit from complementing the focus on learning with *an equally*

strong focus on structure, that is, to address the question: *What are the meta-structures that are decisive to shape meta-learning?*

The reasoning for our proposal derives from the authors' "Argument 3", where they argue that meta-learning makes it easy to manipulate a learning algorithm's complexity to construct resource-rational models of learning. By admitting complexity as an important control for model formation, the authors introduce structural discriminations between meta-learners. But as a scalar measure, complexity cannot avoid "collapsing" qualitatively different structures whenever these are assigned the same complexity. Therefore, we suggest extending the research program beyond scalar orderings as complexity measures: Viewing **meta-structures** as patterns of higher-order structure that are qualitatively different from each other and that offer structural-functional "modules" that can be constructed as entities in their own right and be flexibly used by a meta-learning system. This view draws close inspiration from the advocated neuroscience perspective (their "Argument 4") how constraints of the neurobiological substrate determine the emergence of specific control structures. **Meta-structures** are thus abstractable structural principles guiding the development of "substrate-level" structures in a meta-learning system. While meta-learning summarizes many learning trajectories into an overarching base learner (that can quickly specialize), meta-structure summarizes many learning priors into an overarching "base prior" (that then guides meta-learning efficiently).

Examples for such guiding meta-structures can be found in biological neural network models. As a first meta-structure, we consider **hierarchical organization**: The decomposition of actions into sub-actions on different levels of a hierarchy enables flexible recombination into different behaviors. Hierarchical organization is an established principle of biological motor control that has been applied successfully to Deep Reinforcement Learning (DRL) (Merel, Botvinick, & Wayne, 2019; Neftci & Averbeck, 2019). As a benefit, hierarchical organization enables a form of higher-level learning in which the learner can recombine modular policies into new behaviors without the need to always learn all details from scratch.

As a second example of a meta-structure: **Decentralization** serves parallelization of modules' actions, decoupling of subtasks and factorization of state spaces. Decentralization is well-investigated in motor control in animals, for example, in low-level reflexes, but it is also widely acknowledged that decentralized oscillation generating neuronal circuits are essential for locomotion (cf. Dickinson et al., 2000). While decentralization often is merely characterized as a strategy to cope with slow sensory processing, we emphasize how decentralization facilitates meta-learning. In a study on learning of motor control featuring decentralized modules for a four-legged walker, we showed how decentralization positively affected reinforcement learning on two levels (Schilling, Melnik, Ohl, Ritter, & Hammer, 2021). First, on the basic learning level, decentralization remedies the problem of exponential increase of required training runs in traditional DRL systems as the action space becomes more complex. Decentralization restricts the action space to the much lower number of local actuators, thereby reducing the dimensionality. Without the need for coordination of all control signals by a single centralized controller, the decentralized network learned stable behaviors much faster. Second, on the level of meta-learning, the trained decentralized controller appeared to learn a different, more robust, mapping when compared to a standard centralized controller: The decentralized control structure had learned to transfer previously learned aspects of motor control to entirely new terrains without the need for further context-specific training. Thus, with respect to meta-learning this structural prior (meta-structure) of decentralization proved beneficial for extrapolation of behavior and appears to learn better suited mappings for a broader range of tasks.

A further example of meta-structures can be identified in reversal learning. In **reversal learning**, an agent initially learns a mapping, for example, between certain situations and corresponding appropriate responses, and then finds itself in a situation that *requires a different stimulus response mapping* to achieve behavioral goals. While standard DRL agents learn new mappings at a reversal point from scratch, biological organisms typically solve reversal problems more effectively (Happel et al., 2014): They create already during the initial learning phase hierarchically organized representation structures that can be efficiently used for the new required mapping without learning it from scratch (Jarvers et al., 2016).

The examples imply that a common mechanism by which meta-structures support meta-learning is that of enabling a learning agent to build context. Context-building is naturally introduced by meta-learning itself, for example, as conceived in the target article: An inner-loop learner operates at the fast time scale, with temporally short-ranged contexts, while an "outer-loop process," which could be implemented either as dedicated network modules or as processes made possible on decentralized structures, works at time scales and higher levels of abstraction that allow tuning of inner learner's adaptivity (Schilling, Hammer, Ohl, Ritter, & Wiskott, 2023). This meta-structure can be imagined as recursively extendable, leading to an "onion-like" architecture providing a principled stratification of an overall learning process into a layered hierarchy of learners operating at different levels of granularity (or abstraction), with correspondingly scaled scopes of context. Such concepts can contribute to our understanding of sophisticated learning capabilities as needed by embodied agents that continuously adapt interactions of their body with the environment involving contexts at different temporal and spatial scales.

The discussion of meta-structures illustrates that meta-learning is based on abstractable structural principles that support the generation of a **compositional semantics** linking the functions of modules that emerge from learning in a given context. Meta-structures provide structural preconditions for the establishment of almost arbitrary high-level compositional semantics that can be flexibly reused and serve as building blocks of higher-level abstractions for novel problem solutions without requiring specific training in novel contexts. Together, these aspects underscore the significance of meta-structure for a research program on meta-learning.

Financial support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interests. None.

References

Dickinson, M. H., Farley, C. T., Full, R. J., Koehl, M., Kram, R., & Lehman, S. (2000). How animals move: An integrative view. *Science 288*(5463), 100–106. https://doi.org/10.1126/science.288.5463.100

Happel, M. F., Niekisch, H., Castiblanco Rivera, L. L., Ohl, F. W., Deliano, M., & Frischknecht, R. (2014). Enhanced cognitive flexibility in reversal learning induced by removal of the extracellular matrix in auditory cortex. *Proceedings of the National Academy of Sciences*, *111*(7), 2800–2805.

Jarvers, C., Brosch, T., Brechmann, A., Woldeit, M. L., Schulz, A. L., Ohl, F. W., … Neumann, H. (2016). Reversal learning in humans and gerbils: Dynamic control network facilitates learning. *Frontiers in Neuroscience*, *10*, 535.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Merel, J., Botvinick, M., & Wayne, G. (2019). Hierarchical motor control in mammals and machines. *Nature Communications*, *10*(1), 5489.
Neftci, E. O., & Averbeck, B. B. (2019). Reinforcement learning in artificial and biological systems. *Nature Machine Intelligence*, *1*(3), 133–143.
Schilling, M., Hammer, B., Ohl, F. W., Ritter, H. J., & Wiskott, L. (2023). Modularity in nervous systems – a key to efficient adaptivity for deep reinforcement learning. *Cognitive Computation*, 1–16.
Schilling, M., Melnik, A., Ohl, F. W., Ritter, H. J., & Hammer, B. (2021). Decentralized control and local information for robust and adaptive decentralized deep reinforcement learning. *Neural Networks*, *144*, 699–725.

The added value of affective processes for models of human cognition and learning

Yoann Stussi[a,b]*, Daniel Dukes[a] and David Sander[a,b]

[a]Swiss Center for Affective Sciences, Campus Biotech, University of Geneva, Geneva, Switzerland and [b]Department of Psychology, FPSE, University of Geneva, Geneva, Switzerland
yoann.stussi@unige.ch
daniel.dukes@unige.ch
david.sander@unige.ch
https://www.unige.ch/fapse/e3lab/members1/senior-researchers-and-lecturers/dr-yoann-stussi
https://dukes.space/
https://www.unige.ch/fapse/e3lab/director/

*Corresponding author.

doi:10.1017/S0140525X24000207, e165

Abstract

Building on the affectivism approach, we expand on Binz et al.'s meta-learning research program by highlighting that emotion and other affective phenomena should be key to the modeling of human learning. We illustrate the added value of affective processes for models of learning across multiple domains with a focus on reinforcement learning, knowledge acquisition, and social learning.

Binz et al. bid to establish an ambitious research program to model human cognition using a meta-learning framework. They effectively illustrate the potential and advantages of meta-learned models, showcasing the ability of such models to acquire inductive biases through experience. Notably, the authors outline a compelling blueprint for how these models could foster the development of a domain-general model of human learning. Here, we seek to complement this blueprint by highlighting a key element that is left mostly unexplored in the target article: Affective processes.

Affective processes – typically emotions, feelings, motivations, moods, or attitudes – are not only inherently linked to well-being, but also drive behavioral and cognitive processes such as attention, learning, memory, and decision-making (e.g., LaBar & Cabeza, 2006; Lerner, Li, Valdesolo, & Kassam, 2015; Phelps, 2006; Pool, Brosch, Delplanque, & Sander, 2016). This grants affective processes high explanatory power in understanding human behavior and cognition, a central argument of the affectivism approach (Dukes et al., 2021). As such, we suggest that considering affective processes is pivotal to the modeling of human cognition, and especially of learning. Affective processes are indeed central to – and exert a pervasive influence on – how humans learn (e.g., Öhman & Mineka, 2001; Vollberg & Sander, 2024; Wuensch, Pool, & Sander, 2021). Below, we illustrate how emotion and other affective phenomena are central to human learning across various domains, with a particular focus on reinforcement learning, knowledge acquisition, and social learning.

Affective processes emerge as important factors in reinforcement learning. This fundamental learning process enables individuals to attribute value to states or stimuli and actions via teaching signals such as rewards and punishments. These reinforcers and their associated stimuli typically evoke affective responses, which are core components of reward-seeking and threat-related behaviors (Levy & Schiller, 2021; Stussi & Pool, 2022). Affective processes also modulate how individuals learn from reinforcers. Studies on Pavlovian conditioning – a basic form of reinforcement learning – have shown that stimuli with heightened affective relevance, such as both threat-relevant (e.g., angry faces) and positive emotional (e.g., baby faces) stimuli, are more rapidly and persistently associated with an aversive outcome than neutral stimuli (Stussi, Pourtois, & Sander, 2018; Stussi, Pourtois, Olsson, & Sander, 2021). At the computational level, these studies indicate that affective relevance modulates how individuals learn from prediction errors: Affectively relevant stimuli were associated with a lower learning rate for negative prediction errors (i.e., when the aversive outcome was expected but omitted), enhancing the persistence of their association with the aversive outcome (Stussi et al., 2018, 2021). Similarly, substantial evidence has demonstrated that individuals learn differently about positive and negative outcomes in the instrumental domain (Dorfman, Bhui, Hughes, & Gershman, 2019; Lefebvre, Lebreton, Meyniel, Bourgeois-Gironde, & Palminteri, 2017). Positively valenced prediction errors are generally associated with a higher learning rate than negatively valenced prediction errors, providing a computational correlate of such learning asymmetry (Palminteri & Lebreton, 2022). Altogether, these findings highlight that affective mechanisms shape basic reinforcement learning processes.

Affective processes likewise support epistemic learning. Both positive and negative emotions have long been studied for their roles in the encoding, consolidation, and recall of episodic memories (Levine & Pizarro, 2004), as well as in academic learning in educational settings (see Pekrun & Linnenbrink-Garcia, 2014). Epistemic emotions are the key family of emotions supporting knowledge exploration and acquisition (see Muis, Chevrier, & Singh, 2018). Emotions such as interest, curiosity, confusion, surprise, wonder, or awe are drivers of learning (e.g., Chevrier, Muis, Trevors, Pekrun, & Sinatra, 2019; Vogl, Pekrun, Murayama, & Loderer, 2020). As an illustration of the central role of epistemic emotions in learning, the "trivia questions" paradigm is typically used to understand how epistemic curiosity enhances memory. Using this paradigm, research has shown that the more participants are curious to know the response to a question (e.g., "who is the most cited psychologist of the 21st century?"), the more they later remember the response (e.g., Kang et al., 2009; Marvin & Shohamy, 2016). The impact of curiosity on knowledge exploration and acquisition, partly

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

relying on reward-related processes, is therefore a salient example of how the intrinsic value of information can enhance learning (Murayama, 2022).

While many things can be learned by exploring one's own environment, this individual approach has its limitation. Some information simply cannot be gleaned in this way and requires input from other (human) sources (Harris & Koenig, 2006). Critically, such social learning fundamentally relies on affective processes. Social learning has historically been seen as either a non-human primate phenomenon that explains how behavior can be learned from conspecifics (Zentall & Galef, 1988), or a human cognitive developmental phenomenon concerned with learning from others' testimony (Harris, 2012). However, affective social learning not only points out that these two branches of social learning originate from the same tree (Gruber, Bazhydai, Sievers, Clément, & Dukes, 2022), but also that it is possible to learn from others' affective attitudes about the value of objects (e.g., ideas, people, customs). An important part of who we are – our values, ethics, and morality – is based on our perception, attention, and memory of interaction with and learning from others, whether or not this information is communicated ostensively (Dukes & Clément, 2017; Egyed, Király, & Gergely, 2013). And indeed, what we perceive, attend to, and remember is largely defined by how important, valuable, and affective those objects of perception, attention, and memory are. Not only do we remember what is affectively relevant to us as individuals, but others, serving as proxy relevant detectors, can also signal what is more or less relevant, to be learned or forgotten (Dukes & Clément, 2019; see also Sorce, Emde, Campos, & Klinnert, 1985).

In conclusion, affective processes play a fundamental role in learning across various domains and their consideration is key to the modeling of human learning. Given that emotions are not immutable and static but flexibly arise from the interaction between an individual and their environment (Scherer & Moors, 2019), it could be particularly enlightening to conceptualize emotion as a kind of inductive bias attuned by experience within the meta-learning framework proposed by Binz et al. Such conceptualization could offer a promising way of modeling the effects of emotion on learning, thereby providing added value to the meta-learning research agenda.

Acknowledgements. The authors thank Dr. Eva R. Pool, Professor Dr. Maël Lebreton, and Professor Dr. Fabrice Clément for insightful discussions.

Financial support. Y. S. is supported by an ERC Starting Grant (INFORL-948671) awarded to Professor Dr. Maël Lebreton.

Competing interest. None.

References

Chevrier, M., Muis, K. R., Trevors, G. J., Pekrun, R., & Sinatra, G. M. (2019). Exploring the antecedents and consequences of epistemic emotions. *Learning and Instruction*, *63*, Article 101209. https://doi.org/10.1016/j.learninstruc.2019.05.006

Dorfman, H. M., Bhui, R., Hughes, B. L., & Gershman, S. J. (2019). Causal inference about good and bad outcomes. *Psychological Science*, *30*(4), 516–525. https://doi.org/10.1177/0956797619828724

Dukes, D., Abrams, K., Adolphs, R., Ahmed, M. E., Beatty, A., Berridge, K. C., … … Sander, D. (2021). The rise of affectivism. *Nature Human Behaviour*, *5*, 816–820. https://doi.org/10.1038/s41562-021-01130-8

Dukes, D., & Clément, F. (2017). Author reply: Clarifying the importance of ostensive communication in long-life, affective social learning. *Emotion Review*, *9*(3), 267–269. https://doi.org/10.1177/1754073916679006

Dukes, D., & Clément, F. (Eds.). (2019). *Foundations of affective social learning: Conceptualizing the social transmission of value*. Cambridge University Press. https://doi.org/10.1017/9781108661362

Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological Science*, *24*(7), 1348–1353. https://doi.org/10.1177/0956797612471952

Gruber, T., Bazhydai, M., Sievers, C., Clément, F., & Dukes, D. (2022). The ABC of social learning: Affect, behavior, and cognition. *Psychological Review*, *129*(6), 1296–1318. https://doi.org/10.1037/rev0000311

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Harvard University Press. https://doi.org/10.4159/harvard.9780674065192

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*(3), 505–524. https://doi.org/10.1111/j.1467-8624.2006.00886.x

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhanced memory. *Psychological Science*, *20*(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*, 54–64. https://doi.org/10.1038/nrn1825

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*, Article 0067. https://doi.org/10.1038/s41562-017-0067

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*, 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition*, *22*(5), 530–554. https://doi.org/10.1521/soco.22.5.530.50767

Levy, I., & Schiller, D. (2021). Neural computations of threat. *Trends in Cognitive Sciences*, *25*(2), 151–171. https://doi.org/10.1016/j.tics.2020.11.007

Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, *145*(3), 266–272. https://doi.org/10.1037/xge0000140

Muis, K. R., Chevrier, M., & Singh, C. A. (2018). The role of epistemic emotions in personal epistemology and self-regulated learning. *Educational Psychologist*, *53*(3), 165–184. https://doi.org/10.1080/00461520.2017.1421465

Murayama, K. (2022). A reward-learning framework of knowledge acquisition: An integrated account of curiosity, interest, and intrinsic-extrinsic rewards. *Psychological Review*, *129*(1), 175–198. https://doi.org/10.1037/rev0000349

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037/0033-295X.108.3.483

Palminteri, S., & Lebreton, M. (2022). The computational roots of positivity and confirmation biases in reinforcement learning. *Trends in Cognitive Sciences*, *26*(7), 607–621. https://doi.org/10.1016/j.tics.2022.04.005

Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014). *International handbook of emotion in education*. Routledge.

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, *57*, 27–53. https://doi.org/10.1146/annurev.psych.56.091103.070234

Pool, E., Brosch, T., Delplanque, S., & Sander, D. (2016). Attentional bias for positive emotional stimuli: A meta-analytic investigation. *Psychological Bulletin*, *142*(1), 79–106. https://doi.org/10.1037/bul0000026

Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745. https://doi.org/10.1146/annurev-psych-122216-011854

Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology*, *21*(1), 195–200. https://doi.org/10.1037/0012-1649.21.1.195

Stussi, Y., & Pool, E. R. (2022). Multicomponential affective processes modulating food-seeking behaviors. *Current Opinion in Behavioral Sciences*, *48*, Article 101226. https://doi.org/10.1016/j.cobeha.2022.101226

Stussi, Y., Pourtois, G., Olsson, A., & Sander, D. (2021). Learning biases to angry and happy faces during Pavlovian aversive conditioning. *Emotion*, *21*(4), 742–756. https://doi.org/10.1037/emo0000733

Stussi, Y., Pourtois, G., & Sander, D. (2018). Enhanced Pavlovian aversive conditioning to positive emotional stimuli. *Journal of Experimental Psychology: General*, *147*(6), 905–923. https://doi.org/10.1037/xge0000424

Vogl, E., Pekrun, R., Murayama, K., & Loderer, K. (2020). Surprised-curious-confused: Epistemic emotions and knowledge exploration. *Emotion*, *20*(4), 625–641. https://doi.org/10.1037/emo0000578

Vollberg, M. C., & Sander, D. (2024). Hidden reward: Affect and its prediction errors as windows into subjective value. *Current Directions in Psychological Science*. Advance online publication. https://doi.org/10.1177/09637214231217678

Wuensch, L., Pool, E. R., & Sander, D. (2021). Individual differences in learning positive affective value. *Current Opinion in Behavioral Sciences*, *39*, 19–26. https://doi.org/10.1016/j.cobeha.2020.11.001

Zentall, T. R., & Galef, B. G. (Eds.). (1988). *Social learning: Psychological and biological perspectives*. Psychology Press.

Bayes beyond the predictive distribution

Anna Székely[a,b]* and Gergő Orbán[a]

[a]Department of Computational Sciences, HUN-REN Wigner Research Centre for Physics, Budapest, Hungary and [b]Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Budapest, Hungary
szekely.anna@wigner.hu
http://golab.wigner.mta.hu/people/anna-szekely/
orban.gergo@wigner.mta.hu
http://golab.wigner.mta.hu/people/gergo-orban/

*Corresponding author.

doi:10.1017/S0140525X24000086, e166

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 posterior predictive distributions. In our commentary, we highlight several 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 equivalence 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 are 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 are 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, Chater, Kemp, Perfors, & Tenenbaum, 2010; Kemp, Perfors, & Tenenbaum, 2007; Kemp & Tenenbaum, 2008; Tenenbaum, Griffiths, & Kemp, 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, Meszéna, & Orbán, 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, for example, variational autoencoders (Nagy, Török, & Orbán, 2020; Spens & Burgess, 2024), non-parametric methods (Éltető, Nemeth, Janacsek, & Dayan, 2022; Heald, Lengyel, & Wolpert, 2021; Török et al., 2022), or probabilistic programming (Lake, Salakhutdinov, & Tenenbaum, 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).

Financial support. Supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory.

Competing interest. None.

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.

doi:10.1017/S0140525X24000177, e167

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.

Financial support. No funding to report.

Competing interest. None.

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.

doi:10.1017/S0140525X24000219, e168

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

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

governed by immense numbers of free parameters (up to billions) and are not interpretable in cognitive terms, amounting to a "black box" data-driven approach.

A hyperparameter optimization approach alleviates these concerns by constraining the models it generate to emulate biologically plausible architectures. This allows for formulating and testing mechanistic hypotheses that are based in established literature. The reinforcement meta-learner (RML) model is a good illustration of this framework in the context of executive function (Silvetti, Vassena, Abrahamse, & Verguts, 2018).

Consistent with abundant empirical evidence on biological executive circuits (e.g., Shackman et al., 2011; Silvetti, Seurinck, van Bochove, & Verguts, 2013; Varazzani, San-Galli, Gilardeau, & Bouret, 2015; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011), the RML emulates interactions between the medial prefrontal cortex (MPFC) and two catecholamine nuclei – the ventral tegmental area, releasing dopamine (DA), and the locus coeruleus, releasing norepinephrine (NE). The MPFC module monitors reward rates conveyed by DA and, when detecting a "need for control" (e.g., a decrease in the rates), calls for the release of NE and DA. In turn, these neurotransmitters are broadcast to task-specific cognitive modules and enhance their efficiency, thereby restoring performance and reward rates. The MPFC registers a boost of neurotransmitter release as a cost and uses Bayesian and reinforcement-learning (RL) optimization to learn control settings that maximize rewards while minimizing costs. The RML thus uses traditional Bayesian/RL optimization frameworks to simultaneously regulate motor input and internal cognitive computations, thus modeling both first-order performance and its executive (meta-level) control.

Recent studies have shown that the RML explains empirical findings that have long stumped traditional frameworks, including nonstandard reward modulations in visual areas (Horan, Daddaoua, & Gottlieb, 2019; Silvetti, Lasaponara, Daddaoua, Horan, & Gottlieb, 2023) and curiosity – the intrinsic desire to obtain information in the absence of instrumental rewards (Daddaoua, Lopes, & Gottlieb, 2016; Horan et al., 2019; Silvetti et al., 2023). By monitoring the volatility of the environment, the RML provides a meta-learning-based explanation of the empirical finding of volatility-sensitive learning rates (Silvetti et al., 2013, 2018). Moreover, when coupled to modules emulating memory, motor output, decision making, or attention, the RML reproduces a wide array of behavioral and neural results related, respectively, to memory capacity, motor effort, adaptive regulation of learning rates, and instrumental or curiosity-driven information gathering (Silvetti et al., 2018, 2023). Thus, despite its biologically constrained architecture, the RML gains considerable flexibility and generalizability because it can control different task-specific cognitive computations.

Because the RML uses a biologically plausible architecture with a parsimonious parameter set, it generates a rich set of novel predictions that can be tested against empirical data. These predictions involve possible relationships between behavior and neural activity, between neural activity and neurotransmitter release, and between activity in different brain structures. Existing versions of the RML make predictions about individual computations (e.g., how much memory effort to engage in a particular context) but future versions can be extended to probe how the brain arbitrates between computations (e.g., how it trades-off between relying on memory versus acquiring new sensory information when performing a task).

In conclusion, different meta-learning approaches can differ greatly in their comparative strengths. The entirely unconstrained approach discussed by Binz et al. may be desirable for AI applications where there is no need for biological constraints, for example, when developing an algorithm for a self-driving car, or optimizing planning in multiple tasks. In contrast, we believe that a biologically constrained meta-learning framework is vastly superior for advancing cognitive and neuroscience research (Marblestone, Wayne, & Kording, 2017). Such a biologically constrained framework is grounded in the neuroscientific literature, and can generate testable and falsifiable hypotheses about neurobiological processes underlying cognitive function.

Acknowledgments. Tim Vriens is a PhD student enrolled in the National PhD in Artificial Intelligence, XXXVII cycle, course on Health and life sciences, hosted by Università Campus Bio-Medico di Roma, Italy.

Financial support. M. S. is funded by the Italian Ministry of University and Research, PRIN 2022 program, Grant No. 64.20227MPSEH. M. H. is supported via the Sainsbury Wellcome Centre PhD Programme and has received grants from Reinholdt W. Jorck og Hustrus Fond, Knud Højgaards Fond and Anglo-Danish Society

Competing interests. None.

References

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*, 20–23. https://doi.org/10.1038/538020a

Daddaoua, N., Lopes, M., & Gottlieb, J. (2016). Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. *Scientific Reports*, *6*(1), Article 1. https://doi.org/10.1038/srep20202

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*(4–6), 495–506. https://doi.org/10.1016/s0893-6080(02)00044-8

Horan, M., Daddaoua, N., & Gottlieb, J. (2019). Parietal neurons encode information sampling based on decision uncertainty. *Nature Neuroscience*, *22*(8), 1327–1335. https://doi.org/10.1038/s41593-019-0440-1

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Marblestone, A. H., Wayne, G., & Kording, K. P. (2017). Understand the cogs to understand cognition. *Behavioral and Brain Sciences*, *40*, e272. https://doi.org/10.1017/S0140525X17000218

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain, and cognitive control in the cingulate cortex. *Nature Reviews. Neuroscience*, *12*(3), 154–167. https://doi.org/10.1038/nrn2994

Silvetti, M., Lasaponara, S., Daddaoua, N., Horan, M., & Gottlieb, J. (2023). A reinforcement meta-learning framework of executive function and information demand. *Neural Networks*, *157*, 103–113. https://doi.org/10.1016/j.neunet.2022.10.004

Silvetti, M., Seurinck, R., van Bochove, M., & Verguts, T. (2013). The influence of the noradrenergic system on optimal control of neural plasticity. *Frontiers in Behavioral Neuroscience*, *7*, 160. https://www.frontiersin.org/articles/10.3389/fnbeh.2013.00160

Silvetti, M., Vassena, E., Abrahamse, E., & Verguts, T. (2018). Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. *PLoS Computational Biology*, *14*(8), e1006370. https://doi.org/10.1371/journal.pcbi.1006370

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and dopamine neurons in the reward/effort trade-off: A direct electrophysiological comparison in behaving monkeys. *The Journal of Neuroscience*, *35*(20), 7866–7877. https://doi.org/10.1523/JNEUROSCI.0454-15.2015

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Integrative learning in the lens of meta-learned models of cognition: Impacts on animal and human learning outcomes

Bin Yin*, Xi-Dan Xiao, Xiao-Rui Wu and Rong Lian

School of Psychology, Fujian Normal University, Fuzhou, China.
byin@fjnu.edu.cn
Xiaoxid8899@foxmail.com
Wuxiaorui520@hotmail.com
Lianrong1122@126.com

*Corresponding author.

doi:10.1017/S0140525X2400027X, e169

Abstract

This commentary examines the synergy between meta-learned models of cognition and integrative learning in enhancing animal and human learning outcomes. It highlights three integrative learning modes – holistic integration of parts, top-down reasoning, and generalization with in-depth analysis – and their alignment with meta-learned models of cognition. This convergence promises significant advances in educational practices, artificial intelligence, and cognitive neuroscience, offering a novel perspective on learning and cognition.

Binz et al.'s seminal paper on "Meta-Learned Models of Cognition" offers a transformative view on cognitive modeling, shifting the traditional paradigm toward a more dynamic and experience-based approach. The authors convincingly argue for the superiority of meta-learned models in acquiring inductive biases from experience, as opposed to the rigid, hand-designed structures of traditional models like cognitive architectures and Bayesian models of cognition. This shift represents not only a theoretical advancement but also a practical one, providing a more realistic and adaptable framework for understanding cognitive processes.

Crucially, the paper's synthesis of meta-learning with rational analysis presents an exciting pathway for constructing Bayes-optimal learning algorithms. This approach resonates strongly with integrative learning theories that we have been working on, suggesting a shared trajectory toward developing learning models that can adapt and thrive amidst complexity. Integrative learning refers to the cognitive process of actively integrating learning materials under the influence of metacognition, resulting in an efficient and profound understanding and mastery of knowledge (Lian, 2018). It represents a psychological learning process where metacognition and cognition are highly unified (Yin, Wu, & Lian, 2020, 2023). This learning process model encompasses three modes: holistic integration of parts, top-down reasoning, and generalization for in-depth analysis (Rong Lian, personal communication, March 2020).

Firstly, "holistic integration of parts" involves learners first grasping the overall concept of the subject, establishing a comprehensive initial understanding. This is followed by an exploration of specific parts of the material, with each part being connected back and integrated into this broader understanding, thus reinforcing and enriching the overall comprehension. This "whole-part-whole" learning process has been shown to play a positive role not only in animal learning (Yin et al., 2020, 2023) but also in human learning processes. In studies with university students learning online network knowledge, it was found that, compared to a non-integrative learning group, the integrative learning group better synthesized and processed fragmented online knowledge, resulting in superior learning performance (Huang, 2021). This indicates that individuals, during the learning process, activate metacognition which combines prior knowledge and experience to adjust and optimize new knowledge within working memory, thereby enhancing online learning effectiveness (Mayer, 1997).

Secondly, "top-down reasoning" emphasizes beginning with more broad and generalized high-level concepts and systematically mastering more specific lower-level knowledge points. By understanding and applying higher-level knowledge, they deduce and explain lower-level knowledge, thereby forming a clear, logically structured knowledge framework. Lan (2022) explored the impact of this approach on learning outcomes using a study-recognition paradigm. The results showed that compared to bottom-up learning, top-down reasoning enabled learners to use attention resources more reasonably and effectively, facilitating the relational processing and integration of specific items. Additionally, cognitive semantic processing was smoother, and the integration difficulty between semantics was reduced, significantly enhancing memory effects. Event-related potential (ERP) technology was used to explore the underlying neural mechanisms, revealing that the top-down reasoning group had larger N1 amplitudes and significantly smaller N400 than the bottom-up learning group when learning specific examples. This indicates that top-down reasoning learners more effectively utilized higher-level knowledge, focusing attention resources on more organized processing of specific examples. From the perspective of semantic priming effects, once semantic concepts in memory are activated, their activation can spread to related nodes, increasing their activation level (Meyer & Schvaneveldt, 1971), making the learning of related target stimuli easier (Bueno & Frenck-Mestre, 2002).

Lastly, "generalization for in-depth analysis" starts with learners forming a general representation of the subject, laying the groundwork for the overall framework. Learners then delve into detailed components, engaging in thorough analysis and research. This in-depth study not only deepens understanding of each part but also enhances and refines the initial general framework, culminating in a multifaceted and detailed overall cognition. Chen (2023) explored the impact of this learning mode on university students' reading of texts of varying difficulty. Results showed that the "generalization and in-depth analysis" group had significantly higher understanding and memory scores under different text difficulty conditions compared to the control group, and also had a lower rate of knowledge forgetting. Reading is a process involving simultaneous extraction and construction of meaning (García Madruga et al., 2013). In reading, learners must use complex meaning construction processing to form a complete representation, where cognitive control plays a key role in focusing and switching attention, activating and updating representations (Wu, Tian, Chen, Chen, & Wang, 2021). The "generalization for in-depth analysis" mode helps learners construct complete representations more quickly, and through in-depth analysis

and summarization of new knowledge, continuously adjust and optimize their meaning representations, thereby promoting more refined processing and encoding of information. This process helps learners more effectively retrieve and remember encoded information.

The three modes of integrative learning—holistic integration of parts, top-down reasoning, and generalization with in-depth analysis—closely align with the computational processes of meta-learned models of cognition. This alignment is pivotal, as the initial encounter with a subject in a holistic, higher-level, or generalized manner is essential for setting effective starting metaparameters that guide the learning process. Such an encounter provides a foundational understanding from which learners can refine their perceptions and strategies in a targeted manner. Within this adaptive framework, learners then engage in a cyclical process of interaction, analysis, and metacognitive adjustment, fine-tuning their approach based on this foundational overview and their evolving comprehension of the subject matter. This methodology not only embodies the adaptability characteristic of meta-learning but also supports real-time adjustments during the learning process. As a result, it leads to learning outcomes that are both precisely tailored to the learner's needs and highly effective. Consequently, emphasizing the value of an initial holistic overview reinforces the importance of integrative learning within the meta-learning paradigm. It empowers learners to dynamically adjust their information processing strategies from the outset, significantly enhancing and adapting their learning experiences to achieve optimal outcomes (Rabinowitz, 2019).

The implications of this area of research are vast, offering new directions for educational practices, artificial intelligence, and cognitive neuroscience. In education, these insights could lead to more personalized and effective learning strategies, tailored to individual (meta-)cognitive patterns. For AI, integrating these models could result in more adaptive and intuitive systems, better mimicking human learning processes. In cognitive neuroscience, this research offers potential for deeper understanding of brain-based learning mechanisms. Altogether, this represents a significant stride in our comprehension and application of cognitive and learning sciences, opening new avenues for exploration and innovation.

Financial support. This work was supported by the Humanities and Social Science Funds of the Ministry of Education of China, Project No. 23YJAZH183.

Competing interest. None.

References

Bueno, S., & Frenck-Mestre, C. (2002). Rapid activation of the lexicon: A further investigation with behavioral and computational results. *Brain and Language*, *81*(1–3), 120–130. http://doi.org/10.1006/brln.2001.2511

Chen, S. (2023). *The effect of integrative learning on text reading in elementary and university students* (Master's thesis). Fujian Normal University.

García Madruga, J. A., Elosúa, M. R., Gil, L., Gómez Veiga, I., Vila, J. Ó., Orjales, I., … Duque, G. (2013). Reading comprehension and working memory's executive processes: An intervention study in primary school students. *Reading Research Quarterly*, *48*(2), 155–174. http://doi.org/10.1002/rrq.44

Huang, J. (2021). *The influence of integrative learning on recognition and retrieval of image and text* (Master's thesis). Fujian Normal University. http://doi.org/10.27019/d.cnki.gfjsu.2021.000798

Lan, R. (2022). *The memory advantage effect of top-down reasoning and its cognitive neural mechanism* (Master's thesis). Fujian Normal University. https://doi.org/10.27019/d.cnki.gfjsu.2022.000519

Lian, R. (2018). Integrative learning: Exploring new ways of learning. Fujian Provincial Learning Science Society Symposium, Fuzhou, China.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. http://dx.doi.org/10.1037/h0031564

Rabinowitz, N. C. (2019). Meta-learners' learning dynamics are unlike learner'. arXiv preprint arXiv:1905.01320. https://doi.org/10.48550/arXiv.1905.01320

Wu, S., Tian, L., Chen, J., Chen, G., & Wang, J. (2021). Exploring the cognitive mechanism of irrelevant speech effect in Chinese reading: Evidence from eye movements. *Acta Psychologica Sinica*, *53*(7), 729–745. http://dx.doi.org/10.3724/SP.J.1041.2021.00729

Yin, B., Wu, X., & Lian, R. (2020). An animal behavioral model for the concept of "Integrative Learning". *Acta Psychologica Sinica*, *52*(11), 1278–1287. https://doi.org/10.3724/SP.J.1041.2020.01278

Yin, B., Wu, X.-R., & Lian, R. (2023). "Integrative learning" promotes learning but not memory in older rats. *PeerJ*, *11*, e15101. http://doi.org/10.7717/peerj.15101

Authors' Response

Meta-learning: Data, architecture, and both

Marcel Binz[a,b*], Ishita Dasgupta[c], Akshay Jagadish[a,b], Matthew Botvinick[c], Jane X. Wang[c] and Eric Schulz[a,b]

[a]Max Planck Institute for Biological Cybernetics, Tübingen, Germany; [b]Helmholtz Institute for Human-Centered AI, Munich, Germany and [c]Google DeepMind, London, UK
marcel.binz@helmholtz-munich.de
idg@google.com
akshay.jagadish@helmholtz-munich.de
botvinick@google.com
wangjane@google.com
eric.schulz@helmholtz-munich.de

*Corresponding author.

doi:10.1017/S0140525X24000311, e170

Abstract

We are encouraged by the many positive commentaries on our target article. In this response, we recapitulate some of the points raised and identify synergies between them. We have arranged our response based on the tension between data and architecture that arises in the meta-learning framework. We additionally provide a short discussion that touches upon connections to foundation models.

R1. Introduction

In our target article, we sketched out a research program around the idea of meta-learned models of cognition. The cornerstone of this research program was the observation that neural networks, such as recurrent neural networks, can be trained via meta-learning to mimic Bayesian inference without being explicitly designed to do so (Ortega et al., 2019). This positions the resulting meta-learned models ideally for applications in the context of rational analyses of cognition (Anderson, 2013). Yet, meta-learning additionally enables us to do things that are not possible with other existing methods, thereby pushing the

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

boundaries of rational analyses. Not only is the framework built on solid theoretical grounds, but it also enjoys growing empirical support. Meta-learned models account for a wide range of phenomena that pose a challenge to traditional models, such as the ability for compositional reasoning (Jagadish, Binz, Saanum, Wang, & Schulz, 2023; Lake & Baroni, 2023) or the reliance on heuristic strategies (Binz, Gershman, Schulz, & Endres, 2022; Dasgupta, Schulz, Tenenbaum, & Gershman, 2020).

We believe this research direction is particularly exciting because it allows us to reconceptualize different cognitive processes, including learning, planning, reasoning, and decision-making, into one unified process: The forward dynamics of a deep neural network. In the terminology of modern large language models (LLMs), this ability to acquire knowledge via a simple forward pass is also known as in-context learning (Brown et al., 2020). In-context learning stands in contrast to traditional means of knowledge acquisition in neural networks that requires weight adjustment via gradient descent (hence, it is referred to as in-weights learning). Indeed, there are close connections between meta-learning and training LLMs to which we will return at the end of our response.

Many commentators shared our excitement about this new technology. **McCoy & Griffiths** write that the "direction laid out by Binz et al. is exciting." **Pesnot-Lerousseau & Summerfield** suggest that "this computational approach provides an interesting candidate solution for some of nature's most startling and puzzling behaviours" and that "meta-learning is a general theory of natural intelligence that is – more than [its] classical counterpart – fit for the real world." **Grant** says that "it is an exciting time to be working with and on meta-learning toolkit" but also points out that "many aspects remain open." We agree with this sentiment: Meta-learning is a powerful framework that provides us with the toolkit to build candidate theories of human cognition, but we still must figure out the details and precise instantiations that best describe it.

In the target article, we have framed our argument from a Bayesian angle. Although this offers an invaluable perspective, it does somewhat undermine the role that neural networks play in this context. Indeed, it is really the marriage between Bayesian and neural network models that gives meta-learning its power. There were several commentators who picked up on this. We agree with **Ong, Zhi-Xuan, Tenenbaum, & Goodman (Ong et al.)** who "suggest that the meta-learning approach could be further strengthened by considering connectionist and Bayesian approaches, rather than exclusively one or the other." **McCoy & Griffiths** perhaps put it best by saying that meta-learning "expand[s] the applicability of Bayesian approaches by reconciling them with connectionist models – thereby bringing together two successful research traditions that have often been framed as antagonistic." This integration of research traditions is what enables us to build "constraints from experimental neuroscience, and ecologically relevant environments" into rational theories as suggested by **Grant**, thereby leading to more faithful and naturalistic models.

Many commentators also noted that meta-learning finds applications beyond studying standard human cognition. For example, **Fields & Glazebrook** suggest studying meta-learning "in more tractable experimental systems in which the implementing architecture can be manipulated biochemically and bioelectrically", whereas **Veit & Browning** highlight that "there is also the potential to use meta-learning models to help us understand the evolution of cognition more generally." **Nussenbaum & Hartley** furthermore point out that "these models are particularly useful for testing hypotheses about why learning processes change across development" because they allow us to arbitrate whether changes in an individual are due to an adaption to the external environment (i.e., changes in data) or to internal changes in cognitive capacity (i.e., changes in architecture). We are excited by these research directions as well.

We found the tension between data and architecture laid out by **Nussenbaum & Hartley** very useful, and have therefore decided to organize our response around it. We begin by discussing the commentaries that placed a focus on the importance of data for understanding human cognition (sect. R2), followed by those that focused on the importance of model architecture instead (sect. R3). The point where those two concepts meet will be the centerpiece of our discussion (sect. R4). We finish our response by clarifying some of the misunderstandings that have arisen from our original target article (sect. R5), before providing a general conclusion (sect. R6).

R2. Data matters more than we thought

Historically, cognitive models have been largely based on symbolic representations. Examples include models of heuristic decision-making, problem-solving, or planning. This modeling tradition is largely based on the premise that model architectures are the driving factor in determining behavior. Proponents of this approach often argue that symbolic representations are necessary to capture core ingredients of human cognition, such as decision-making, problem-solving, or planning (Marcus, 1998). The advent of Bayesian models of cognition expanded on this picture. Even though most Bayesian models are also based on symbolic representations, they are sensitive to the data that are expected to be encountered. If assumptions about the environment change, the behavior of these models changes as well. The past 30 years have shown that people are indeed adaptive to the environment, thereby providing considerable support to Bayesian models of cognition (Griffiths, Kemp, & Tenenbaum, 2008).

In contrast to models with symbolic representations, neural networks are based on distributed vector representations. Many have argued that neural networks are inherently ill-equipped for reasoning, planning, and problem-solving because they lack the symbolic representations of their cousins. Indeed, there is a whole line of research (known as neurosymbolic AI) attempting to fix these issues by incorporating symbolic processes into neural network architectures (De Raedt, Manhaeve, Dumancic, Demeester, & Kimmig, 2019). The framework of meta-learning demonstrates that this may not be necessary. It instead offers a proof-of-concept showing that – when trained on the right data – neural networks can exhibit many emergent phenomena that have traditionally been attributed to symbolic models, such as the ability for model-based (Wang et al., 2016) and compositional reasoning (Lake & Baroni, 2023). For example, as already discussed in our target article, Lake and Baroni (2023) have shown that a vanilla transformer architecture can be taught to make compositional inferences via meta-learning. Findings like this allow us to interpret human compositionality as "an emergent property of an inner-loop, in-context learning algorithm that is itself meta-learned" as discussed by **Russin, McGrath, Pavlick, & Frank**. Likewise, Wang et al. (2016) have shown that a simple meta-learned recurrent neural network can act like a model-based reinforcement learning algorithm, even though it does not contain any explicit architecture that facilitates model-based

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

reasoning. The implications of these findings are vast as pointed out by **Pesnot-Lerousseau & Summerfield** who suggest that "many supposedly 'model-based' behaviours may be better explained by meta-learning than by classical models" and that meta-learning "invites us to revisit our neural theories of problem solving and goal-directed planning."

Taken together, this suggests that model architecture may not be as important as once thought for building systems with human-like reasoning capabilities. What matters much more than we initially thought however are the data these systems are trained on. If the data are generated by symbolic processes, meta-learning will pick up on this and compile these processes into the resulting models.

There is evidence from recent work in NeuroAI supporting the idea that data trumps architecture. In a large-scale analysis, Conwell, Prince, Kay, Alvarez, and Konkle (2022), for example, found different model architectures achieve near equivalent degrees of brain predictivity in the human ventral visual system and that the data they were trained on had a much bigger influence. Muttenthaler, Dippel, Linhardt, Vandermeulen, and Kornblith (2022) presented similar findings, suggesting that "model scale and architecture have essentially no effect on the alignment [between the representations learned by neural networks and] human behavioral responses, whereas the training dataset and objective function both have a much larger impact."

That puts the focus on the question of "what to learn?" **Prat & Lamm** argue that this is the hard problem in natural and artificial intelligence. They further point out that nature solves this problem via evolution and that we cannot handcraft the utility function (or error measurements) for each task separately. We sympathize with this perspective. However, the evolutionary perspective is not the most useful one when the goal is to build models of human cognition. We certainly do not want to simulate the entire process of evolution for this. If we want to avoid this, what can we do as an alternative? For one, we can use automated tools, such as Gibbs sampling with people (Harrison et al., 2020), that measure the priors and utility functions of people and plug the resulting data into our pipelines. That this is possible in the meta-learning framework has been recently demonstrated by Kumar et al. (2022). There is also recent work suggesting that the generation of data reflecting the real world can be automated using foundation models. Jagadish, Coda-Forno, Thalmann, Schulz, and Binz (2024) have, for example, shown that this is a promising approach for building models that can acquire human-like priors when trained on ecologically valid problems. In particular, they queried a LLM to generate naturalistic classification problems, trained a meta-learning system on these problems, and demonstrated that the resulting meta-learned models explain many effects observed in the literature.

R3. Yet, architecture matters too

However, it is certainly not only data that matter for understanding human cognition. Model architecture will still play a role as pointed out in several commentaries (e.g., **Schilling, Ritter, & Ohl**). In fact, there are already results showing that this is the case in the meta-learning setting. For example, Chan et al. (2022) studied the trade-off between in-context and in-weights learning and found that in-context learning only emerges when the training data exhibit certain data distributional properties. Importantly, this was only true in transformer-based models but not in recurrent models (which relied on in-weights learning instead). This demonstrates that different model architectures can lead to characteristically different behaviors, thereby highlighting that architecture *is* crucial – at least to some extent. From a cognitive perspective, the interesting question will be how much architecture is needed.

Many commentaries suggested that enhancing the black-box meta-learning framework with process-level structures will help us to better understand human cognition. We agree that this is an intriguing line of thought (see the discussion in our target article within sects. 2.4 and 5). In many cases, the commentaries added an additional dimension to our original proposal. We discuss some of these proposals in the following and place them into the context of our framework.

Sanborn, Yan, & Tsvetkov (Sanborn et al.) highlight that people often deviate from normative behavior (whether Bayesian or meta-learned). Earlier work has shown that many of these deviations can be captured by rational process models, which approximate posterior predictive distributions using posterior mean, posterior median, or other summary statistics. These rational process models play hand-in-hand with the meta-learning framework as pointed out by Sanborn et al.. Essentially, their proposal is to have a rational process model reason based on the meta-learned posterior predictive distribution. This combination brings together the best of both worlds as one does not even have to retrain the meta-learning model as in other approaches that induce limited computational resources into meta-learned models (Binz & Schulz, 2022; Saanum, Éltető, Dayan, Binz, & Schulz, 2023). That can be convenient from a practical perspective. We agree that this is an appealing property and look forward to how the interaction between these two frameworks will play out in the future.

In a similar vein, **Grant** argues that the meta-learning toolkit needs stronger architectural constraints. Her proposal emphasizes a connectionist implementation of meta-learning called model-agnostic meta-learning (MAML). MAML implements its stepwise updating using gradient descent as opposed to the models we focused on in our target article whose updating is implemented using neural network forward dynamics (which is also referred to as memory-based meta-learning). Although MAML involves meta-learning with the same objective as discussed in our target article, it differs in what is being meta-learned. In MAML, one adapts the initial weights of a neural network, such that subsequent gradient steps lead to optimal learning. That leads to an interesting class of gradient-based meta-learned models that have many (but not all) of the advantages discussed in our target article. MAML's key feature is that it allows for a seamless link between the algorithmic and the computational level of analysis. Future research should compare different classes of models against each other and find the one that best explains human behavior. It will be particularly exciting to pit gradient-based models (such as MAML) against memory-based models (including recurrent neural networks) and see which class of theories offers a better account of human behavior. Doing so will allow us to answer some of the big, outstanding questions of cognitive psychology and neuroscience, such as whether we can find any evidence for computations like gradient descent and backpropagation in the brain.

Last but not least, **Cea** and **Stussi, Dukes, & Sander** suggest that the meta-learning framework could benefit from the inclusion of affective elements. We agree that doing so can provide added value to the meta-learning research agenda. Yet, at the same time, this proposal highlights one of the tensions involved

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

in building complex systems, namely deciding on what should be prewired and what should be given the chance to emerge instead. To illustrate this point, let us consider one of the examples provided by Stussi et al. highlighting the importance of affective processes: For humans, positively valenced prediction errors are generally associated with a higher learning rate than negatively valenced prediction errors (Palminteri & Lebreton, 2022). In a recent study, we found that this characteristic emerges naturally in meta-learned models (Schubert, Jagadish, Binz, & Schulz, 2024), thereby illustrating that at least some affective processes are already present in meta-learned models.

Ultimately, determining which inductive biases should be prewired and which should be learned from data depends on which research question one is investigating. If one wants to obtain a process-level understanding of a phenomenon, there is no better way than formalizing that phenomenon mathematically and simulating it in silico. If the goal, on the other hand, is to simply induce superhuman-like general abilities in a computational model, modern machine learning research, such as the work on AlphaZero, has taught us that we should keep the amount of prewiring limited and instead rely mainly on the data itself (Sutton, 2019).

R4. Transcending levels of analysis

The full power of meta-learning does not solely come from its close ties to Bayesian inference – the algorithmic implementation also matters. To get a more complete understanding of human cognition, it seems likely that we need to consider both data and architecture. Meta-learning allows us to do this by bringing together two modeling traditions that have focused on these two aspects individually. It combines the advantages of Bayesian models – which feature powerful, data-dependent inductive biases – and neural network models – which come with a vast amount of architectural design choices – seamlessly. To quote Ortega (2020), meta-learning "brings back Bayesian statistics within deep learning without even trying – no latents, no special architecture, no special cost function, nada."

That was also recognized by some of the commentators. **McCoy & Griffiths** state that meta-learning "reconcil[es Bayesian approaches] with connectionist models – thereby bringing together two successful research traditions that have often been framed as antagonistic". This feature allows the framework to effortlessly jump between different levels of analysis, from the computational over the algorithmic to the implementational. Furthermore, although neural networks have been often criticized for not being able to engage in symbolic reasoning, meta-learning illustrates that it is, in principle, possible to equip neural networks with symbolic inductive biases.

Nussenbaum & Hartley highlight potential applications in the context of developmental psychology. Here, one of the central questions involves identifying whether "age-related changes in learning reflect adaptation to age-varying 'external' ecological problems *or* 'internal' changes in cognitive capacity." We believe that this is an exciting research direction. In fact, we have recently applied some of these ideas to test whether the developmental trajectories of children in the context of intuitive physics can be captured with deep generative models by manipulating the amount of training data or the system's computational resources (Buschoff, Schulz, & Binz, 2023).

However, the strict dichotomy between data and architecture is likely to be false. Instead, the two interact with each other over the lifespan, as also pointed out by **Nussenbaum & Hartley**. Meta-learning allows us to disentangle the two and study them jointly or separately. This not only has implications for understanding adults and kids but also in the context of mental and physical health. We can, for instance, ask which kinds of environments cause or exacerbate certain mental illnesses, or what types of architectural constraints lead to maladaptive behaviors. In doing so, we might be able to better understand these issues and, in turn, potentially develop targeted aids for them.

R5. Points of contention

Although the framework proposed in our target article was received well overall, there were a few points of contention raised by some of the commentators. In this section, we address and clarify these issues.

The first of them is raised by **Ong et al.** and by **Székely & Orbán**. They both argue that having to specify an inference problem is a virtue of the Bayesian approach, not a limitation. From their perspective, the process of defining the inference problem can in itself shed light on the system whose cognitive processes are being modeled. We agree that this is a valid – and often very useful – strategy. However, both of the commentaries come with the implicit assumption that this is not possible in the meta-learning framework. We think this is a wrong dichotomy. The exact same research strategy can be applied in the meta-learning framework: (1) Define a data-generating distribution, (2) draw samples from it, (3) use these to construct a meta-learned model, and (4) compare models with different assumptions against each other. To illustrate this using the probabilistic programming example of Ong et al., one could, for example, define a distribution over probabilistic programs and use meta-learning to construct a neural network that can perform approximate inference over probabilistic programs. Although we generally agree that this is a useful research strategy, it is important to mention that there are settings in which it is just not applicable, as outlined in our target article. We think this is where the strength of meta-learning lies. It allows us to do all of the things we can do in the traditional Bayesian framework – including probabilistic programs – and more.

From a conceptual perspective, we also have to weigh in on the commentaries of **Vriens, Horan, Gottlieb, & Silvetti** who state that "the framework generates models that are not interpretable in cognitive terms and, and crucially, are governed by an immense numbers of free parameters […]." That is not an accurate depiction of the meta-learning framework from our perspective. Although these models have potentially a lot of parameters, they are not free parameters that are fitted to human data. Instead, they are only optimized to maximize performance on a given task. Vriens et al. further claim that the framework generates models with low falsifiability. This is also far from the truth. Every meta-learned model can be compared against alternative models, and hence potentially falsified, as we and others have shown in many previous studies. Indeed, this is true not only on the behavioral level but also on the neural level (i.e., when there are inconsistencies with neural recordings). In this sense, meta-learned models provide even stronger grounds on which they can be refuted as pointed out by **Grant**. The meta-learning framework as a whole is of course harder to falsify. However, we believe that it should be the role of a framework to generate useful theories, not to be falsifiable.

We also noticed that there were a few misunderstandings concerning the distinction between tool and theory highlighted in our

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

target article. In particular, we put forward the notion of using meta-learning as a tool for building models of human learning. We did not say much or make any claims about how people actually acquire their learning algorithms, i.e., the process of meta-learning itself (see sect. 1.4 of the target article). Although this is an important problem, it is outside the scope of our article. **Llewellyn** states that our first question is to understand how people improve their learning abilities over time (and subsequently that we fail to address this question satisfyingly). However, we explicitly want to highlight again that we did not strive to address this question in our target article. Likewise, **Calderan & Visalli** mention that we should position ourselves against hierarchical Bayesian models, which can be used to model learning-to-learn. We agree that this would be needed if we were targeting to find out how people learn-to-learn, which we are not. Even though the study of learning-to-learn is outside the scope of our target article, we still believe that the meta-learning framework could provide an interesting perspective for studying these processes. This was, for example, noted by **Nussenbaum & Hartley** in the context of developmental psychology, or by **Yin, Xiao, Wu, & Lian** who make the connection to integrative learning.

Finally, **Calderan & Visalli** question the utility of meta-learning to build rational models in large worlds. In particular, they ask: "What justifications exist for the selection of training data?" They rightfully claim that meta-learned models have priors too and that they therefore offer no important advantages over Bayesian models. However, in contrast to Bayesian models, meta-learned models do not require an explicit expression for these priors – they only need samples from them, which is a much weaker requirement. That means that we can go out and measure them by collecting samples. In turn, this gives us many opportunities. We can, for example, ask people to generate samples from their priors (as done in the work of Kumar et al. [2022] mentioned earlier), or we can go out and determine priors that match real-world statistics (as done in the work of Jagadish et al. [2024] mentioned earlier). Meta-learning then allows us to compile these priors into a computational model. Although hierarchical Bayesian models may also be able to construct their priors, as mentioned by Calderan & Visalli, they can only do so in a predetermined class of functions, preventing an effective application to large world problems.

R6. Links to foundation models

To our surprise, none of the commentaries touched upon the similarities between meta-learning and training LLMs. We therefore wanted to use this opportunity to raise a few points on this topic ourselves. Essentially, LLMs are trained using the same objective we have discussed in our target article (equation 7). The only thing that is special is that the data distribution amounts to the whole internet. In this sense, LLMs can be viewed as a special case of meta-learned models – all the same principles apply. Thus, one way to view LLMs is that they approximate Bayesian inference to predict the next tokens in human language. Like the meta-learned models we have discussed in our target article, LLMs learn from their context (i.e., a history of previous observations) to make better predictions with more examples by updating only their internal activations. There are exciting research questions only waiting to be answered in the space between human cognition and LLMs, and we believe that the meta-learning perspective could help us in this endeavor (Binz & Schulz, 2023; Hussain, Binz, Mata, & Wulff, 2023; Yax, Anlló, & Palminteri, 2023).

It might also be interesting to think about meta-learned models that are not based on the objectives outlined in our target article. For example, we may ask how meta-learned models relate to the concepts of free energy minimization and active inference (see commentary by **Penacchio & Clemente**), what objectives are needed to meta-learn quantum models (see commentaries by **Clark** and **Mastrogiorgio**), whether we can meta-learn models using contrastive losses (Tian et al., 2020), or whether it is possible to give meta-learning systems the ability to determine their own objectives (see commentary by **Moldoveanu**). Doing so might lead to models that do not approximate Bayesian inference but have other appealing properties. Nevertheless, putting theoretical properties aside, finding out which models are most useful in understanding cognition will ultimately be an empirical question, not a theoretical one.

Where will cognitive modeling be in 10 years from now? We predict that there will be major advances in two main directions: (1) Our models will become much more domain-general and (2) they will process high-dimensional, naturalistic stimuli. The meta-learning framework will help us to achieve both of these objectives. The first is already addressed by design: Meta-learning involves training on a collection of tasks – we only have to make this collection more diverse. Regarding the second, meta-learned models of cognition can be readily combined with visual neural networks, thereby giving them the ability to "see" experimental stimuli similarly to people (as pointed out by **Sanborn et al.**). We are already witnessing some of these systems that perform a wide range of tasks in complex, vision-based environments in the machine learning literature. Examples include models such as Voyager (Wang et al., 2023), Ada (Team et al., 2023), or SIMA (Raad, Ahuja, Besse, Bolt, & Young, 2024) – all of which are based (at least to some extent) on a meta-learned model. Unfortunately, these models are currently too expensive to train for most academic research labs (let alone to run ablations on them). For example, training Ada requires access to 64 TPUs for five weeks. However, compute is getting cheaper every year, and – together with technological advances – we think it is likely that a similar system could be trained on standard hardware 10 years from now. We are excited by this prospect and what it means for understanding human cognition.

References

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

Binz, M., & Schulz, E. (2022). Modeling human exploration through resource-rational reinforcement learning. *Advances in Neural Information Processing Systems*, *35*, 31755–31768.

Binz, M., & Schulz, E. (2023). Turning large language models into cognitive models. *arXiv preprint arXiv:2306.03917*.

Binz, M., Gershman, S. J., Schulz, E., & Endres, D. (2022). Heuristics from bounded meta-learned inference. *Psychological Review*, *129*(5), 1042.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Buschoff, L. M. S., Schulz, E., & Binz, M. (2023). The acquisition of physical knowledge in generative neural networks. In *International Conference on Machine Learning* (pp. 30321–30341). PMLR.

Chan, S., Santoro, A., Lampinen, A., Wang, J., Singh, A., Richemond, P., … Hill, F. (2022). Data distributional properties drive emergent in-context learning in transformers. *Advances in Neural Information Processing Systems*, *35*, 18878–18891.

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., & Konkle, T. (2022). What can 1.8 billion regressions tell us about the pressures shaping high-level visual representation in brains and machines?. *BioRxiv*, 2022-03.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

Dasgupta, I., Schulz, E., Tenenbaum, J. B., & Gershman, S. J. (2020). A theory of learning to infer. *Psychological Review*, *127*(3), 412.

De Raedt, L., Manhaeve, R., Dumancic, S., Demeester, T., & Kimmig, A. (2019). Neuro-symbolic= neural+ logical+ probabilistic. In *NeSy'19@ IJCAI, the 14th International Workshop on Neural-Symbolic Learning and Reasoning*.

Griffiths, T. L., Kemp, C., Tenenbaum, J. B.. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). Cambridge University Press.

Harrison, P., Marjieh, R., Adolfi, F., van Rijn, P., Anglada-Tort, M., Tchernichovski, O., … Jacoby, N. (2020). Gibbs sampling with people. *Advances in Neural Information Processing Systems*, *33*, 10659–10671.

Hussain, Z., Binz, M., Mata, R., & Wulff, D. U. (2023). A tutorial on open-source large language models for behavioral science. *PsyArXiv preprint*.

Jagadish, A. K., Binz, M., Saanum, T., Wang, J. X., & Schulz, E. (2023). Zero-shot compositional reinforcement learning in humans.

Jagadish, A. K., Coda-Forno, J., Thalmann, M., Schulz, E., & Binz, M. (2024). Ecologically rational meta-learned inference explains human category learning. *arXiv preprint arXiv*:2402.01821.

Kumar, S., Correa, C. G., Dasgupta, I., Marjieh, R., Hu, M. Y., Hawkins, R., … Griffiths, T. (2022). Using natural language and program abstractions to instill human inductive biases in machines. *Advances in Neural Information Processing Systems*, *35*, 167–180.

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*(7985), 115–121.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, *37*(3), 243–282.

Muttenthaler, L., Dippel, J., Linhardt, L., Vandermeulen, R. A., & Kornblith, S. (2022). Human alignment of neural network representations. *arXiv preprint arXiv:2211.01201*.

Ortega, P. A. (2020). Twitter. Retrieved from https://twitter.com/adaptiveagents/status/1334609817432940550?lang=bn

Ortega, P. A., Wang, J. X., Rowland, M., Genewein, T., Kurth-Nelson, Z., Pascanu, R., … Legg, S. (2019). Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030*.

Palminteri, S., & Lebreton, M. (2022). The computational roots of positivity and confirmation biases in reinforcement learning. *Trends in Cognitive Sciences*, *26*(7), 607–621.

Saanum, T., Éltető, N., Dayan, P., Binz, M., & Schulz, E. (2023). Reinforcement learning with simple sequence priors. *Advances in Neural Information Processing Systems*, *36*, 61985–62005.

Schubert, J. A., Jagadish, A. K., Binz, M., & Schulz, E. (2024). In-context learning agents are asymmetric belief updaters. *arXiv preprint arXiv:2402.03969*.

SIMA Team, Raad, M. A., Ahuja, A., Barros, C., Besse, F., Bolt, A., … Young, N. (2024). Scaling instructable agents across many simulated worlds. Technical Report.

Sutton, R. (2019). The bitter lesson. *Incomplete Ideas (blog)*, *13*(1), 38.

Team, A. A., Bauer, J., Baumli, K., Baveja, S., Behbahani, F., Bhoopchand, A., … Zhang, L. (2023). Human-timescale adaptation in an open-ended task space. *arXiv preprint arXiv:2301.07608*.

Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., & Isola, P. (2020). What makes for good views for contrastive learning? *Advances in Neural Information Processing Systems*, *33*, 6827–6839.

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., … Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.

Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., … Botvinick, M. (2016). Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*.

Yax, N., Anlló, H., & Palminteri, S. (2023). Studying and improving reasoning in humans and machines. *arXiv preprint arXiv:2309.12485*.

https://doi.org/10.1017/S0140525X23003266 Published online by Cambridge University Press

2.2 In-context learning agents are asymmetric belief updaters

Schubert, J. A., **Jagadish, A. K.**, Binz, M.*, & Schulz, E.* (2024). In-Context Learning Agents Are Asymmetric Belief Updaters. In Forty-first International Conference on Machine Learning (ICML). doi:10.5555/3692070.3693859. arXiv:2402.03969.

Contributions in-context

Optimism bias in a reinforcement learning setting is the tendency of an agent to learn more from positive reward prediction errors than from negative ones [123–125]. Although this asymmetric belief updating has been well established in humans, it remained unclear whether large language models (LLMs) exhibit similar behavior and, more importantly, why would any agent, natural or artificial, acquire such a tendency.

In this work, we first subjected LLMs to reinforcement learning tasks in-context and analyzed their behavior using the Rescorla-Wagner model and its variants. We found that, like humans, LLMs also exhibited asymmetric belief updating when learning from valenced rewards [126]. Surprisingly, similar to humans, this tendency (1) reverses in LLMS when learning from counterfactual feedback, that is, when learning from rewards of unchosen options; (2) disappears when no agency is implied, such as when rewards are passively observed without any control over action selection (see also Section 3.1 in Background).

To understand the rationale behind these effects, we derived idealized in-context reinforcement learning algorithms via meta-learning on a task distribution whose characteristics matched those on which LLMs and humans were evaluated; see Section 1.3 in Outlook and Section 2.1 in Publications for details. Specifically, we meta-learned on bandit tasks varying in reward distribution, feedback type, and control. The resulting models allowed us to perform a rational analysis of the exhibited behavior; see Section 1.6 in Background. Upon analyzing their learning dynamics, we found that meta-learned algorithms replicated all three behavioral patterns observed in humans and LLMs. These findings indicated that the asymmetric belief updating in reinforcement learning settings emerges as a rational strategy, shaped by optimal adaptation to the structure of the environments previously encountered by the agent.

We concluded by reflecting on the external validity of these findings, speculating on methods to investigate the mechanistic underpinnings of optimism bias, whether it will persist in settings where asymmetric belief updating is suboptimal by design, and the role of computational cognitive models in studying internal mechanisms underlying LLM behavior (see Section 3.1 in Outlook for an extended discussion).

In-Context Learning Agents Are Asymmetric Belief Updaters

Johannes A. Schubert [1] **Akshay K. Jagadish** [1 2] **Marcel Binz** [* 1 2] **Eric Schulz** [* 1 2]

Abstract

We study the in-context learning dynamics of large language models (LLMs) using three instrumental learning tasks adapted from cognitive psychology. We find that LLMs update their beliefs in an asymmetric manner and learn more from better-than-expected outcomes than from worse-than-expected ones. Furthermore, we show that this effect reverses when learning about counterfactual feedback and disappears when no agency is implied. We corroborate these findings by investigating idealized in-context learning agents derived through meta-reinforcement learning, where we observe similar patterns. Taken together, our results contribute to our understanding of how in-context learning works by highlighting that the framing of a problem significantly influences how learning occurs, a phenomenon also observed in human cognition.

1. Introduction

Large language models (LLMs) are powerful artificial systems that excel at many tasks (Radford et al., 2019). They can, among other things, write code (Roziere et al., 2023), help to translate from one language to another (Kocmi & Federmann, 2023), and play computer games (Wang et al., 2023). Their abilities are so far-reaching that some (Bubeck et al., 2023) have argued that they "could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system". At the same time, they are notoriously difficult to interpret which becomes especially aggravating as these models permeate through our society.

In the present paper, we aim to shed light on the in-context learning abilities of LLMs (Brown et al., 2020). Making use of the two-alternative forced choice (2AFC; Fechner, 1860) paradigm from cognitive science, we show that in-context learning implements an asymmetric updating rule when learning about the values of options. In particular, we find that – when provided with outcomes from freely chosen options – in-context learning exhibits an optimism bias (Sharot, 2011), meaning that it learns more from positive than from negative prediction errors. We additionally find that this effect is mediated by two factors. First, when the outcome of the unchosen option is also provided, the bias for that option reverses and the model learns more from negative than from positive prediction errors. Furthermore, when no agency is implied and the query says *someone else* does the sampling instead of *you* sampled, the bias disappears. Interestingly, similar behavioral effects have been observed in human subjects (Lefebvre et al., 2017; Chambon et al., 2020).

Why do these tendencies for asymmetric belief updating emerge in both natural and artificial agents? Previous work has suggested that asymmetric belief updating might be a rational strategy to implement as it allows agents to achieve maximum rewards in a given task. However, these claims have been limited by the use of a restricted model class (Lefebvre et al., 2022; Cazé & van der Meer, 2013). To investigate this idea further, we study the behavior of idealized in-context learning agents trained specifically to solve 2AFC tasks using meta-reinforcement learning (Meta-RL). Meta-RL agents have been shown to implement Bayes-optimal learning strategies upon convergence (Ortega et al., 2019; Binz et al., 2023b) and enable us to test if displaying such a bias is rational. We again find the same behavioral effects in these agents: (1) they show an optimism bias when only observing outcomes of the chosen option, (2) the bias reverses when learning about the value of the unchosen option, and (3) it disappears when the agent has no control about its own choices.

Taken together, our results have broad implications for both natural and artificial agents. We have shown that in-context learning depends critically on how the problem is framed. There are many applications where practitioners have control over problem framing, and thus our results suggest that these design choices must be carefully considered to achieve desired outcomes. In the context of human cognition, our

[*]Equal contribution [1]Computational Principles of Intelligence Lab, Max Planck Institute for Biological Cybernetics, Tübingen, Germany [2]Institute for Human-Centered AI, Helmholtz Computational Health Center, Munich, Germany. Correspondence to: Johannes A. Schubert <johannes.schubert@tue.mpg.de>.

Proceedings of the 4 [s] International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

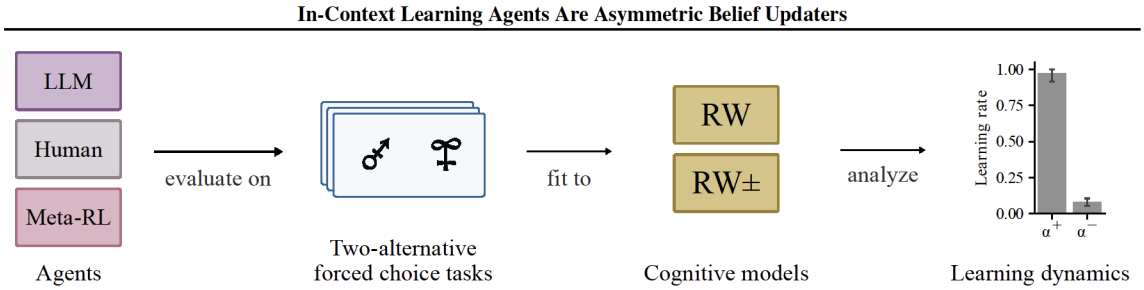

Figure 1. Schematic of our methodology, where we evaluate the learning dynamics of LLMs, humans, and meta-reinforcement learning (Meta-RL) agents on two-alternative forced choice tasks. After evaluating the agents on the tasks, we fit variants of cognitive models based on the Rescorla-Wagner (RW) model to the resulting behavior. Finally, we analyze the fitted models and extract and compare the learning rates.

simulations extend previous work suggesting that the optimism bias (and related effects) may not be a bias after all, as it can be considered a rational response to certain problems.

2. Related work

In-context learning in LLMs. In their seminal paper introducing GPT-3, Brown et al. (2020) demonstrated that LLMs can do tasks, such as text translation, question-answering, and arithmetic problems, after seeing just a few demonstrations – an ability they referred to as in-context learning. There has been a recent surge of research trying to better understand this phenomenon: investigating when and how in-context learning emerges (Chan et al., 2022; Min et al., 2022; Wei et al., 2023), identifying what algorithms LLMs implement during in-context learning (Xie et al., 2021; Von Oswald et al., 2023), and mapping out what can be learned in-context (Garg et al., 2022; Dong et al., 2022). For the present paper, the work of Binz & Schulz (2023) and Coda-Forno et al. (2023) is of particular relevance. They showed that LLMs can learn to perform simple multi-armed bandit problems – which were similar to the 2AFC problems used here – in-context and without requiring any weight updates.

Asymmertric belief updating in cognitive science. How humans integrate task-relevant information to update their beliefs has received a lot of attention in cognitive science (Jacobs & Kruschke, 2011; Nassar et al., 2010; Gershman, 2015). Traditionally, this question has been investigated using the 2AFC paradigm, which provides a controlled setup to study the various facets of human reinforcement learning, including generalization, exploration, and compositional inference (Jagadish et al., 2023; Binz & Schulz, 2022; Lefebvre et al., 2017; Behrens et al., 2007; Chambon et al., 2020; Palminteri & Lebreton, 2022; Schulz et al., 2019).

The experimental paradigms and analyses used in this paper are heavily inspired by the following studies. Lefebvre et al. (2017) showed that people have asymmetric belief updating tendencies in a reinforcement learning setting. Later on, Chambon et al. (2020) demonstrated that this tendency reverses when people observe outcomes for unchosen options and that it completely disappears in purely observational trials (i.e. when participants observe outcomes following a predetermined choice). Recently, Palminteri & Lebreton (2022) reviewed the influence of factors, such as reward magnitude (Lefebvre et al., 2022) and volatility (Gagne et al., 2020; Behrens et al., 2007), on asymmetric belief updating in people. Finally, Lefebvre et al. (2022) used simulations to show that asymmetric belief updating can lead to optimal performance under certain reward regimes. In a series of experiments, the authors manipulated the reward probabilities of options. They found that an optimism bias is optimal for low reward probabilities, while a pessimism bias is optimal for high reward probabilities. This occurs as the learning asymmetry enables better separation of learned reward probabilities in their respective reward regimes and decreases the probability of switching to the worse option after a certain amount of trials (Cazé & van der Meer, 2013).

3. Methods

In this section, we first describe how we queried an LLM to perform 2AFC tasks and explain how models from cognitive psychology can be used to analyze the in-context learning dynamics of LLMs and idealized agents derived through Meta-RL. We provide an overview of our methodology in Figure 1.

3.1. LLM prompting

In a 2AFC task, an agent has to repeatedly choose between two options and receives a reward after each choice. The goal of the agent is to maximize the reward over all trials. We prompted an LLM to perform such 2AFC tasks. We used Claude-1.2 as the reference LLM for all our experiments

via its API[1] with the temperature set to 0.0.

The prompt design was based on earlier work that had studied LLMs in similar settings (Binz & Schulz, 2023; Coda-Forno et al., 2023). Each prompt included an introduction to the task setup, a history of previous observations, and a question asking for the next choice. The tasks were framed in a gambling context, where an agent visits several casinos with two slot machines. More details on the prompting are provided in Appendix A.1.

The following prompt was used to run the default 2AFC task (see Section 4):

Prompt Task 1

You are going to visit four different casinos (named 1, 2, 3, and 4) 24 times each. Each casino owns two slot machines which all return either 0.5 or 0 dollars stochastically with different reward probabilities. Your goal is to maximize the sum of received dollars within 96 visits.

You have received the following amount of dollars when playing in the past:

- Machine B in Casino 4 delivered 0.5 dollars.
- Machine F in Casino 1 delivered 0.0 dollars.
- Machine B in Casino 4 delivered 0.5 dollars.

Q: You are now in visit 4 playing in Casino 4. Which machine do you choose between Machine R and Machine B?

A: Machine [insert]

Prompts were updated dynamically after every trial. Slot machines were labeled with a random letter, excluding meaningful ones (U, I), and the order of slot machine labels was randomized. The selected slot machine returned a stochastic reward that was appended to the bulleted history of previous slot machine interactions in subsequent prompts. We simulated the behavior of the LLM for a certain number of trials and recorded the action-reward pairs.

We additionally considered two other variants of 2AFC tasks, which shared the same general structure but differed in how information was conveyed to the agent. One provided additional information by revealing the reward for the unchosen option (see Section 5) and the other included trials in which the choice of a particular option was forced (see Section 6).

[1] https://www.anthropic.com/news/introducing-claude

3.2. Using cognitive models to analyze in-context learning dynamics

For the analysis of the learning dynamics of an agent, we built simplified but interpretable cognitive models of its choice behavior. We then identified which parameter setting in these models provides the best explanation for the observed data. The resulting parameter values can then offer a window into the behavior of the agent. This approach has its origins in cognitive science (Miller, 1956; Rescorla & Wagner, 1972; Wilson & Collins, 2019) but has recently been adopted to study the behavior of artificial agents as well (Dasgupta et al., 2022; Binz & Schulz, 2023; Bigelow et al., 2023).

The core of our analysis is the Rescorla-Wagner (RW) model (Rescorla & Wagner, 1972) – a classic model for studying learning in humans [2]. It formalizes learning as a dynamic process of minimizing prediction errors between expected and actual outcomes:

$$V_{t+1}(a) = V_t(a) + \alpha \cdot \delta_t$$
$$\delta_t = r_t - V_t(a)$$

where δ_t is the prediction error between the observed reward r_t and the expected reward value $V_t(a)$ in trial t. The learning rate α determines how much the expected value for action a changes after an observation. Thus, it quantifies the amount of learning that occurs based on the prediction error. The model maps learned values to choice probabilities using a softmax decision rule with an inverse temperature parameter β:

$$p(a) = \frac{\exp(\beta \cdot V_t(a))}{\sum_{k=1}^{K} \exp(\beta \cdot V_t(k))}$$

There are several extensions to the RW model that have been used to evaluate different hypotheses about how learning occurs. We relied on one such extension, namely the RW$\pm$ model (Palminteri & Lebreton, 2022), which introduces separate learning rates for positive and negative prediction errors:

$$V_{t+1}(a) = V_t(a) + \begin{cases} \alpha^+ \cdot \delta_t, & \text{if } \delta_t > 0 \\ \alpha^- \cdot \delta_t, & \text{if } \delta_t < 0 \end{cases}$$
$$\delta_t = r_t - V_t(a)$$

Positive prediction errors occur in situations where the received reward is greater than the estimated value (i.e. $\delta_t > 0$), while negative prediction errors occur when the received reward is less than the estimated value (i.e. $\delta_t < 0$).

[2] Note that while the RW model is considered one of the most canonical modeling choices for 2AFC tasks, other alternatives have been used to model human behavior in this setting, such as Bayesian models (Zhang & Yu, 2013; Gershman, 2018) and drift-diffusion models (Pedersen et al., 2017; Lefebvre et al., 2022).

This model allows us to study how the information of prediction errors is weighted during learning. In the case where the expected value is influenced symmetrically by both prediction errors (i.e. $\alpha^+ = \alpha^-$), the RW$\pm$ model is equivalent to the classical RW model. If the learning rates differ, the beliefs are updated asymmetrically, either optimistically ($\alpha^+ > \alpha^-$) or pessimistically ($\alpha^- > \alpha^+$).

We relied on a maximum a posteriori estimation approach to fit the parameters of these models to the behavioral data generated by an LLM. For the RW model, this involves fitting parameters $\theta = [\alpha, \beta]$ (two parameters in total). For the RW$\pm$ model, this involves fitting parameters $\theta = [\alpha^+, \alpha^-, \beta]$ (three parameters in total). Prior probabilities were based on Daw et al. (2011), using a beta distribution $\mathcal{B}(1.1, 1.1)$ for learning rates and a gamma distribution with $\mathcal{G}(1.2, 5)$ for the inverse temperature parameter. Each option was represented by a separate expected value V_t and the initial values V_0 were assigned to the average reward values. More details on the parameter estimation can be found in Appendix A.2.

For each task, we compared different cognitive models based on their average posterior probabilities (PP; Wu et al., 2020). To do this, we first computed the Bayesian Information Criterion (BIC; Schwarz, 1978) for each model m and each simulation:

$$\text{BIC}_m = k \cdot \ln(N) - 2 \cdot \ln p(a_{1:N}|\hat{\theta}_m)$$

where $\hat{\theta}$ are the estimated parameters based on the agent's choices, k is the number of estimated parameters, and N is the number of choices performed. Under the assumption of a uniform prior over models, the PP for each simulation can then be approximated as:

$$p(m|a_{1:N}) \approx \frac{\exp(-0.5 \cdot \text{BIC}_m)}{\sum_i \exp(-0.5 \cdot \text{BIC}_i)}$$

The BIC is a standard metric for model comparison and selection. It can be interpreted as an approximation to the model evidence (or marginal likelihood) which is obtained by integrating out the parameters of a model using Laplace's method (Bishop, 2006).

4. Partial information results in optimism bias

We started our investigations with the 2AFC task with default settings in which only the outcome for the chosen slot machine was revealed as illustrated in the prompt of the previous section and in Figure 2a. We adopted the experimental paradigm of an earlier study with humans from Lefebvre et al. (2017). It involved four different casinos with win probabilities of 0.25/0.25, 0.25/0.75, 0.75/0.25, and 0.75/0.75 for the respective slot machines. Winning led to a reward of 0.5 dollars, losing to a reward of 0.0 dollars. Each casino was visited 12 times in a random order resulting in 48 visits in total. We simulated the LLM 50 times on the task.

First, we examined performance regarding regret, which is the amount of reward missed relative to the optimal choice. When averaged across all simulations and casinos, regret decreased significantly from 0.08 ± 0.01 in the first trial to 0.05 ± 0.01 in the last trial ($t(49) = 3.1, p = .002$; see Figure 2b). This difference between starting and final performance during in-context learning is similar to that observed in human studies (Lefebvre et al., 2017).

To investigate whether in-context learning is done symmetrically or asymmetrically, we fitted the classical RW and the RW$\pm$ model to the simulated behavior of the LLM separately for each simulated run as described in Section 3.2. The model comparison indicated that the RW$\pm$ model provided on average a better explanation of the data with a PP of 0.91 ± 0.08 (see Figure 2c). Analyzing the learning rates revealed that this was associated with a strong optimistic asymmetry: new information about the options was incorporated more readily when it was desirable (positive prediction errors) than undesirable (negative prediction errors) as shown in Figure 2d, with $\alpha^+ > \alpha^-$ ($\alpha^+ = 0.68 \pm 0.02$, $\alpha^- = 0.08 \pm 0.03$; $t(49) = 22.2, p < 0.001$). In-context learning seems to overweigh new evidence that conveys a positive valence. Previous work with human subjects has observed a similar – although less pronounced – optimism bias (reproduced in Figure 2d for reference).

We additionally tested how well our findings generalize to different LLMs and task formats. We therefore repeated our experiment on seven LLMs, including Claude-2.1, Claude-3 Haiku, GPT-4, Llama-2-7B, Llama-2-7B-Chat, Llama-2-70B, and Llama-2-70B-Chat. Furthermore, we extended our analysis to include a broader range of task formats. In all of these settings, we found robust evidence for asymmetric belief updating. The results of these analyses are presented in Appendix B.

5. Pessimistic updating for unchosen options

Next, we investigated the in-context learning dynamics of LLMs in a setting that provided full feedback about the outcomes of both the chosen and unchosen slot machines. For this, we borrowed another experimental paradigm of an earlier study with humans from Chambon et al. (2020). The adapted task consisted of visits to multiple casinos, each containing two slot machines. Each of the casinos was prompted independently of the others. Half of the casinos provided full feedback, the other half only provided partial

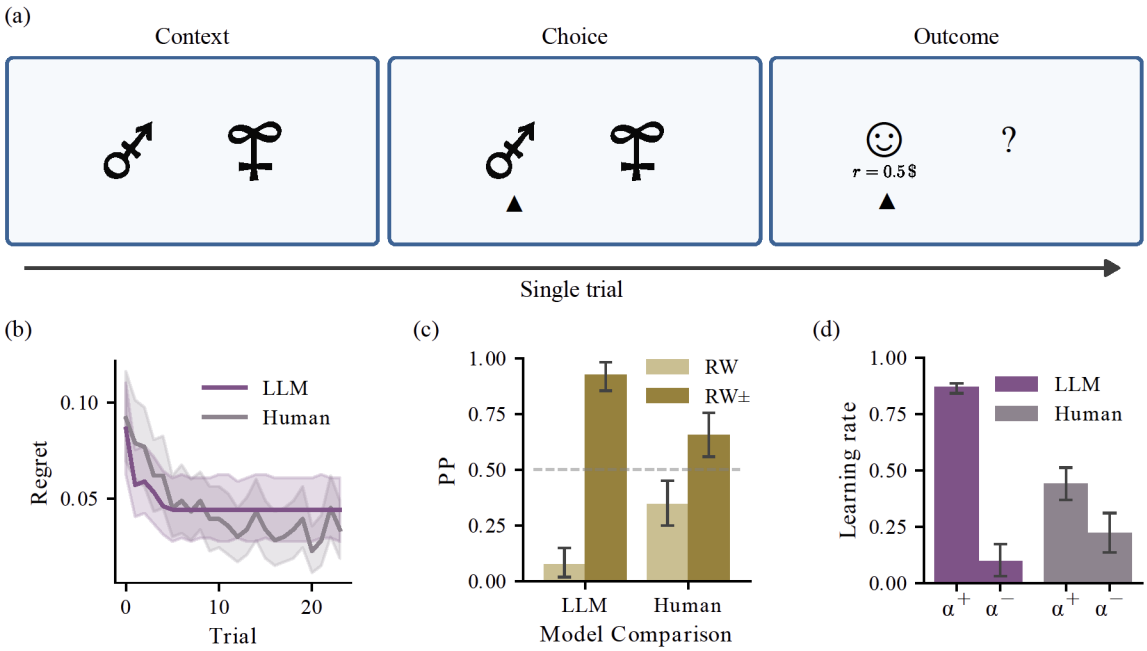

Figure 2. 2AFC task with partial feedback. (a) Presentation of a single trial. First, two slot machines, shown as symbols, are presented. After a choice is made, the outcome is shown. (b) Average performance of the LLM and humans measured in terms of regret. Performance improves over trials. (c) Model comparison of the Rescorla-Wagner (RW) model and the RW± model. For both the LLM (left) and human participants (right), RW± provides a better fit to the data, as indicated by the average posterior probability (PP). (d) Average learning rates of the RW± model for the LLM and human participants. Both agents show a stronger response to positive prediction errors than to negative prediction errors. Human participant data reproduced from (Lefebvre et al., 2017). Error bars and shaded areas and correspond to 95% CIs.

feedback about the chosen machine.[3] In the full feedback casinos, the foregone outcome of the unchosen slot machine was provided in addition to the outcome of the chosen slot machine in the history of prior interactions (see Figure 3a). We adapted the feedback items as follows to reflect this change: "On visit 1 you played Machine H and earned 1.0 point. On Machine E you would have earned -1.0 point."

Half of the casinos were high-reward casinos with reward probabilities of 0.9 and 0.6 for both machines, and the other half were low-reward casinos with reward probabilities of 0.4 and 0.1. In all casinos, the outcome was either a gain or a loss of one point. The prompt in the full feedback casinos also mentioned that the observed outcome of the unchosen machine would not be added to the total rewards earned (see Appendix C.1). We simulated the LLM 24 times on the task consisting of 16 casinos with 40 trials each.

When comparing the partial to the full feedback casinos, we saw a decrease in final regret from partial feedback (0.07 ± 0.01) to full feedback casinos (0.02 ± 0.01), with $t(191) = 2.9, p = .004$ (see Figure 3b). This suggests that in-context learning was able to incorporate the additional information conveyed by the unchosen option. Like in the previous section, performance in terms of regret was comparable to that observed in an earlier study with human subjects as shown in Figure 3c.

For analyzing the dynamics of in-context learning, we extended the RW± model to also account for the additional information provided by the unchosen option. This extended RW± model included separate learning rates for positive and negative prediction errors of the chosen and unchosen options (i.e. four learning rates in total).

Figure 3d illustrates that fitted learning rates for chosen and unchosen options show opposite asymmetric patterns. While we again observed an optimism bias for learning about the chosen machine ($t(23) = 5.7, p < .0001$), information for the unchosen machine was integrated such that negative prediction errors were preferentially taken into account, relative to positive ones ($t(23) = 5.7, p < .0001$).

[3] Note that the task also contained forced-choice trials in which the agent has to select a predetermined machine. We ignored these trials for the analysis presented in this section, but come back to them in the next section.

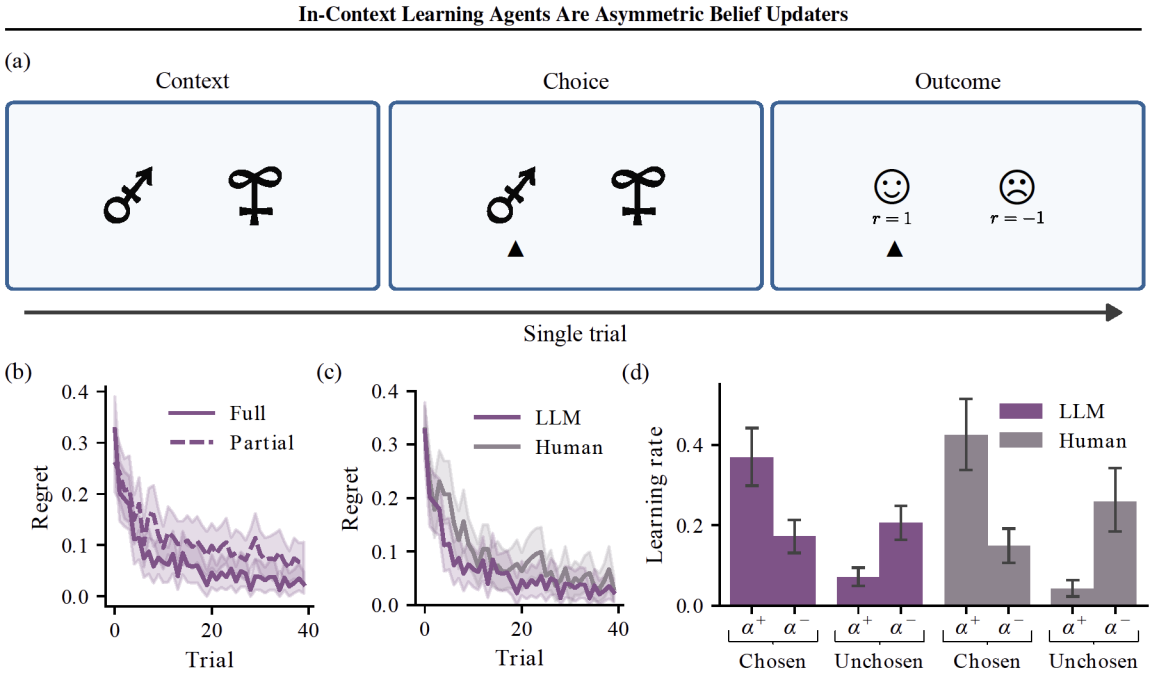

Figure 3. 2AFC task with full feedback. (a) Presentation of a single trial: The two slow machines are again shown as symbols. After a choice is made, the outcome of both the chosen and the unchosen option is shown. (b) Average regret of the LLM in partial and full feedback casinos, showing that the additional information of full feedback casinos leads to improved performance. (c) Average regret of the LLM and humans for full feedback casinos. The performance of the LLM improves over trials and is on par with human performance. (d) Learning rates of the full feedback model with two learning rates – for positive and negative prediction errors – for the chosen and unchosen slot machine. Both agents have an optimism bias for the chosen option and a pessimism bias for the unchosen option. Human participant data reproduced from (Chambon et al., 2020). Error bars and shaded areas correspond to 95% CIs.

This pattern is known as confirmation bias as it refers to integrating information in a way that confirms prior beliefs (Palminteri et al., 2017; Nickerson, 1998). Earlier studies with people in a matched experiment (Chambon et al., 2020) found similar behavioral characteristics (see Figure 3d for reference).

The pattern of learning rates suggests that the four learning rates can be compressed into just two: a confirmatory learning rate that combines the positive chosen and negative unchosen learning rates and a disconfirmatory learning rate that combines the negative chosen and positive unchosen learning rates. We therefore fitted a second cognitive model that was obtained by merging the learning rates depending on the confirmatory or disconfirmatory outcome. This simplified model provided a better fit to the data with an average PP of 0.93 ± 0.03 (refer to Figure 8 for the learning rates of this model). This implies that LLMs update their beliefs about certain outcomes more when new evidence confirms their prior beliefs and past decisions than when it disconfirms or contradicts them.

6. No asymmetric updating without agency

In our final analysis, we examined the influence of agency on the in-context learning dynamics of LLMs by providing additional information about slot machines through observational trials. We used the same general task structure as in the previous section (Chambon et al., 2020). However, half of the visited casinos now included randomly interleaved forced-choice trials of another agent playing in the casino (mixed-choice casinos; see Figure 4a). The other half contained only free-choice trials to assess the performance improvements resulting from the additional information provided by the forced-choice trials in the mixed-choice casinos. Both types of casinos provided partial feedback about the outcome of the selected machine.

We adapted the prompt structure of the force-choice trials as follows: “On visit 1 someone else played Machine H and received 1.0 point”. To avoid biasing the agent towards a particular machine, the forced-choice trials sampled both slot machines equally. The prompt of the mixed-choice casinos mentioned that rewards from forced-choice trials would not be added to the total reward (see Appendix D.1). As in the previous section, half of the casinos were high-

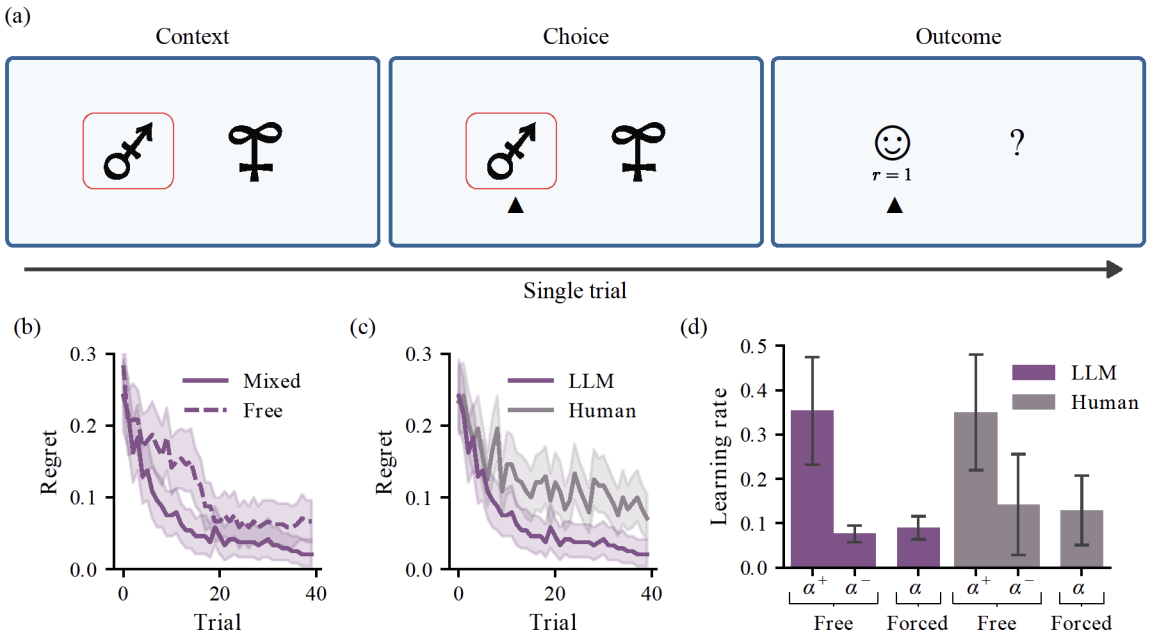

Figure 4. 2AFC task for the agency condition (a) Presentation of a single forced-choice trial: In forced-choice trials, one of the two slot machines is preselected (red square) and its outcome is presented directly to the LLM. (b) Average regret of the LLM in mixed-choice and free-choice casinos, showing that the additional information of forced-choice trials leads to improved performance in mixed-choice casinos. (c) Average LLM and human regret for mixed-choice casinos. The performance of the LLM outperforms the human participants over trials. (d) Learning rates of the 3α model with two learning rates for the free-choice trials and one learning rate for the forced-choice trials. Both agents integrate feedback for positive and negative prediction errors in free-choice trials asymmetrically, whereas feedback from forced-choice trials is integrated symmetrically. Human participant data reproduced from (Chambon et al., 2020). Error bars and shaded areas correspond to 95% CIs.

reward casinos, the other half were low-reward casinos. The six mixed-choice casinos consisted of 80 trials with 40 free and 40 forced-choice trials, while the six free-choice casinos consisted of only 40 free-choice trials. We simulated the LLM on this task 24 times.

When comparing the free-choice casinos with the mixed-choice casinos, we found that the LLM incorporated the additional information from the forced-choice trials, leading to performance improvements with a decrease in regret from 0.07 ± 0.2 to 0.02 ± 0.01 ($t(143) = 2.5, p = 0.01$; see Figure 4b). In comparison to human data from an earlier study (Chambon et al., 2020), the LLM learned significantly faster in this setting as shown in Figure 4c (final regret for LLMs: 0.02 ± 0.01, final regret for humans: 0.07 ± 0.02; $t(143) = 2.7, p = .008$).

To analyze the effects of choice types on learning we fitted two different cognitive models to the behavior of the LLM in the mixed-choice casinos: (1) a 4α model which consisted of separate learning rates for positive and negative prediction errors for both free-choice and forced-choice trials and (2) a 3α model which consisted of separate learning rates for positive and negative prediction errors for free-choice trials and only one learning rate for forced-choice trials. The model comparison indicated that the 3α model represented the behavior of the LLM best ($PP_{3\alpha} = 0.78 \pm 0.06$; see Figure 4c), suggesting that it seems to integrate the information from forced-choice trials symmetrically (i.e. $\alpha^+ = \alpha^-$). In contrast, information from free-choice trials is asymmetrically weighted with $\alpha^+ = 0.36 \pm 0.06$ being greater than $\alpha^- = 0.08 \pm 0.01$ ($t(23) = 4.7, p = .0001$), as seen in Figure 4d. This implies the extent to which the LLM can control its environment changes how it integrates received information.

7. Idealized in-context learning agents also display asymmetric updating

To better understand why in-context learning exhibits these behavioral characteristics, we tested whether they also emerge in a more controlled setting. For this, we trained idealized transformer-based agents to solve our previously examined 2AFC tasks via in-context learning. The agent

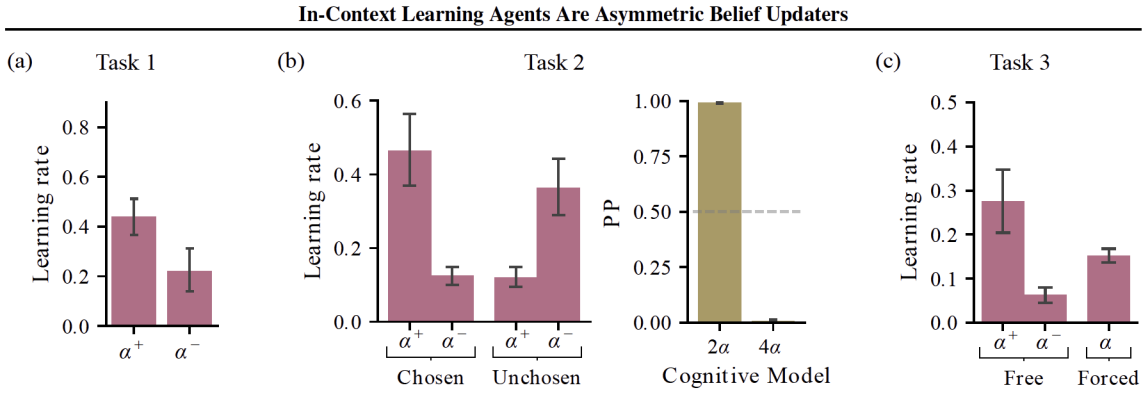

Figure 5. Learning rate analyses for the Meta-RL agent. (a) In the partial feedback task, the RW± provided a better fit to the Meta-RL agents' behavior ($PP_{RW\pm} = 0.97 \pm 0.02$) and showed an optimistic tendency to integrate information. (b) In the full feedback task, the tendency to integrate positive outcomes for chosen options optimistically and negative outcomes for unchosen options pessimistically is even more pronounced than in the LLMs (left). Model comparison showed that the simplified confirmatory model (2α) fits the data better (right). (c) In the agency condition, the 3α best fit the simulated behavior ($PP_{3\alpha} = 0.85 \pm 0.04$), implying that information was integrated asymmetrically in free-choice trials and symmetrically in forced-choice trials. Error bars correspond to 95% CIs.

received the previously selected action a_{t-1} and the reward that followed r_{t-1}, alongside the current context c_t which varied slightly for each task (see Appendix E.1), as inputs in each time-step. We trained the agent using Meta-RL (Duan et al., 2016; Wang et al., 2016) to learn a history-dependent policy $\pi_{\boldsymbol{\Theta}}$ that maximizes the expected sum of rewards over a prespecified task distribution:

$$\max_{\boldsymbol{\Theta}} \mathbb{E}_{p(\theta, c_{1:T}) \prod p(r_t|a_t,c_t,\theta)\pi_{\boldsymbol{\Theta}}(a_t|a_{1:t-1},r_{1:t-1},c_{1:t})} \left[\sum_{t=1}^{T} r_t \right]$$

where $a_{1:t-1}$, $r_{1:t-1}$, and $c_{1:t}$ denote sequences of actions, rewards, and contexts respectively, while $\boldsymbol{\Theta}$ denotes the weights of the underlying transformer. The task distribution used for training is specified by $p(\theta, c_{1:T})$ and $p(r_t|a_t, c_t, \theta)$ with θ corresponding to a vector containing win probabilities for all slot machines.

The distribution shared a similar structure in all three experiments. We assumed that win probabilities for each option were sampled from a uniform distribution at the beginning of each training episode to capture the assumption of an uninformative prior. The agent consisted of a Transformer network (Vaswani et al., 2017) with a model dimension of $8 \cdot \text{input\_size}$, two feedforward layers with a dimension of 128, and eight attention heads, followed by two linear layers that output a policy and a value estimate, respectively. The network weights were adjusted by gradient-based optimization using ADAM (Kingma & Ba, 2014) on a standard actor-critic loss (Mnih et al., 2016) at the end of each training episode. Further details about the training are provided in Appendix E.

We trained all Meta-RL agents until convergence and then tested their in-context learning abilities without performing any further updates to the network weights. It has been shown that the history-dependent policy learned in this setting can solve new but similar tasks in an approximately Bayes-optimal way (Mikulik et al., 2020; Ortega et al., 2019). The intuition behind this result is that the Meta-RL protocol incentivizes the agent to maximize a Monte-Carlo approximation of the Bayes-optimal objective. While analytical solutions to this objective are intractable for most cases, Meta-RL allows us to derive a tractable approximation, and thereby to investigate belief updating in an idealized setting.

We simulated the agents on all three tasks and analyzed their behavior as described in the previous sections. We found that the idealized transformer-based agents learned strategies that outperformed both LLMs and humans as shown in Figure 11. Furthermore, the idealized agents showed similar learning characteristics to LLMs: (1) in the partial feedback task, they learned more from positive prediction errors, (2) the pattern reversed for the counterfactual option in the case of the full feedback task, and (3) asymmetric updating was limited to free-choice trials and was absent in forced-choice trials. These results are illustrated in Figure 5.

Taken together, these results indicate that behavior observed in humans and LLM shares key characteristics with idealized in-context learning agents trained specifically on 2AFC tasks.

8. Discussion

People change their learning strategies based on how the problem is framed (Palminteri & Lebreton, 2022). In this paper, we have shown that this also holds for in-context learning agents. In particular, we found that LLMs exhibit

an optimism bias, i.e. they learn more from better-than-expected outcomes (positive prediction errors) than from worse-than-expected ones (negative prediction errors). However, this bias was only present when the prompt was formulated in a way that implied agency. Furthermore, we found that for counterfactual feedback for unchosen options, the bias reversed and the model learned more from negative than positive errors for these options.

We conducted these analyses in a highly controlled setting, providing high internal validity to our results. However, claims regarding their external validity must be taken with care for now, and future studies will have to investigate whether we can also find similar patterns in more naturalistic settings. Furthermore, our findings relied on an inference from observed learning rate patterns in cognitive models. It is unclear if the link is causal as another auxiliary computational process could potentially be explaining the observed pattern. Nevertheless, opposing interpretations of the pattern have not yet been substantiated by experimental evidence in human experiments (Palminteri & Lebreton, 2022). To test the causal relationship between the latent processes underlying cognitive models and those of the LLM, we see two possibilities. The first is to lesion the neurons in the LLM that encode trial-by-trial negative prediction errors to determine if it results in a stronger optimistic bias. The second is to swap the activations between positive and negative prediction error encoding neurons to determine if it results in a pessimism bias instead of an optimism bias.

In the tasks we have investigated, it is rational to perform asymmetric belief updating – as indicated by our simulations with Meta-RL agents. It remains an open question if this also holds in situations where asymmetric belief updating is suboptimal. Future work should aim to characterize whether in-context learning also displays asymmetric belief updating in such situations. The study of Globig et al. (2021) could be a starting point for this question as it provides an example of a setting where people show an optimism bias even though it is not rational.

From a methodological perspective, our work demonstrates that it is possible to fit simpler computational models to the behavior of LLMs and use the resulting parameter values to infer *how* they behave. Fitting and interpreting parameters of simpler computational methods provides us with a tool that complements existing techniques for explainable machine learning (Roscher et al., 2020). We believe that there are further exciting applications in this research field.

Taken together, our results contribute to our understanding of how in-context learning in LLMs works, which is especially important as the number of applications of these models in real-world scenarios is increasing (Binz et al., 2023a; Eloundou et al., 2023; Kasneci et al., 2023). If the biases found in this work also emerge in tasks where they are not optimal, as has been shown in humans (Shepperd et al., 2013), it will be important to develop techniques to mitigate them.

Acknowledgements

We thank the members of the "Computational Principles of Intelligence Laboratory" (CPI Lab), and Stefano Palminteri for their comments, discussions, and support. We would also like to thank the authors of Lefebvre et al. (2017), and Chambon et al. (2020) for making the data from their study available. Further, we thank all reviewers for their constructive feedback that helped to improve our work. This work was supported by the Max Planck Society, the Volkswagen Foundation, and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645.15/18.

Impact statement

Large language models (LLMs) excel in a wide range of applications due to their in-context learning abilities. Our research explores in-context learning and therefore contributes to our understanding of these models.

We have demonstrated certain behavioral patterns emerge depending on how a task is framed. This knowledge can be used for good and bad. On the good side, we now better understand how LLMs behave in certain situations. That means we can take precautionary measures if we see divergences from desired behaviors. However, bad actors may also exploit the fact that LLMs have an optimism bias, for example in the context of disinformation campaigns, risk assessments, or user recommendations. There may be other potential societal consequences of our work that we are not aware of.

References

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9):1214–1221, September 2007. ISSN 1546-1726. doi: 10.1038/nn1954.

Bigelow, E. J., Lubana, E. S., Dick, R. P., Tanaka, H., and Ullman, T. D. In-context learning dynamics with random binary sequences. *arXiv preprint arXiv:2310.17639*, 2023.

Binz, M. and Schulz, E. Modeling human exploration through resource-rational reinforcement learning. *Advances in Neural Information Processing Systems*, 35: 31755–31768, 2022.

Binz, M. and Schulz, E. Using cognitive psychology to understand gpt-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120, 2023.

Binz, M., Alaniz, S., Roskies, A., Aczel, B., Bergstrom, C. T., Allen, C., Schad, D., Wulff, D., West, J. D., Zhang, Q., et al. How should the advent of large language models affect the practice of science? *arXiv preprint arXiv:2312.03759*, 2023a.

Binz, M., Dasgupta, I., Jagadish, A., Botvinick, M., Wang, J. X., and Schulz, E. Meta-Learned Models of Cognition, April 2023b.

Bishop, C. M. Pattern recognition and machine learning. *Springer*, 2:5–43, 2006.

Botvinick, M., Ritter, S., Wang, J., Kurth-Nelson, Z., Blundell, C., and Hassabis, D. Reinforcement Learning, Fast and Slow. *Trends in Cognitive Sciences*, 23, April 2019. doi: 10.1016/j.tics.2019.02.006.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901, 2020.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023.

Cazé, R. D. and van der Meer, M. A. Adaptive properties of differential learning rates for positive and negative outcomes. *Biological cybernetics*, 107(6):711–719, 2013.

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., and Palminteri, S. Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, 4(10):1067–1079, October 2020. ISSN 2397-3374. doi: 10.1038/s41562-020-0919-5.

Chan, S., Santoro, A., Lampinen, A., Wang, J., Singh, A., Richemond, P., McClelland, J., and Hill, F. Data distributional properties drive emergent in-context learning in transformers. *Advances in Neural Information Processing Systems*, 35:18878–18891, 2022.

Coda-Forno, J., Binz, M., Akata, Z., Botvinick, M., Wang, J. X., and Schulz, E. Meta-in-context learning in large language models. *arXiv preprint arXiv:2305.12907*, 2023.

Dasgupta, I., Lampinen, A. K., Chan, S. C., Creswell, A., Kumaran, D., McClelland, J. L., and Hill, F. Language models show human-like content effects on reasoning. *arXiv preprint arXiv:2207.07051*, 2022.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215, 2011.

Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., Sun, X., Xu, J., and Sui, Z. A survey for in-context learning. *arXiv preprint arXiv:2301.00234*, 2022.

Duan, Y., Schulman, J., Chen, X., Bartlett, P. L., Sutskever, I., and Abbeel, P. Rl $\hat{2}$: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*, 2016.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*, 2023.

Fechner, G. T. *Elemente der Psychophysik*, volume 2. Breitkopf u. Härtel, 1860.

Gagne, C., Zika, O., Dayan, P., and Bishop, S. J. Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife*, 9:e61387, December 2020. ISSN 2050-084X. doi: 10.7554/eLife.61387.

Garg, S., Tsipras, D., Liang, P. S., and Valiant, G. What can transformers learn in-context? a case study of simple function classes. *Advances in Neural Information Processing Systems*, 35:30583–30598, 2022.

Gershman, S. J. A unifying probabilistic view of associative learning. *PLoS computational biology*, 11(11):e1004567, 2015.

Gershman, S. J. Deconstructing the human algorithms for exploration. *Cognition*, 173:34–42, 2018.

Globig, L. K., Witte, K., Feng, G., and Sharot, T. Under threat, weaker evidence is required to reach undesirable conclusions. *Journal of Neuroscience*, 41(30):6502–6510, 2021.

Jacobs, R. A. and Kruschke, J. K. Bayesian learning theory applied to human cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(1):8–21, 2011.

Jagadish, A. K., Binz, M., Saanum, T., Wang, J. X., and Schulz, E. Zero-shot compositional reinforcement learning in humans, July 2023.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103: 102274, 2023.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kocmi, T. and Federmann, C. Large language models are state-of-the-art evaluators of translation quality. *arXiv preprint arXiv:2302.14520*, 2023.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., and Palminteri, S. Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, 1(4):0067, March 2017. ISSN 2397-3374. doi: 10.1038/s41562-017-0067.

Lefebvre, G., Summerfield, C., and Bogacz, R. A Normative Account of Confirmation Bias During Reinforcement Learning. *Neural Computation*, 34(2):307–337, January 2022. ISSN 0899-7667, 1530-888X. doi: 10.1162/neco_a_01455.

Mikulik, V., Delétang, G., McGrath, T., Genewein, T., Martic, M., Legg, S., and Ortega, P. Meta-trained agents implement bayes-optimal agents. *Advances in neural information processing systems*, 33:18691–18703, 2020.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

Min, S., Lyu, X., Holtzman, A., Artetxe, M., Lewis, M., Hajishirzi, H., and Zettlemoyer, L. Rethinking the role of demonstrations: What makes in-context learning work? *arXiv preprint arXiv:2202.12837*, 2022.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. Asynchronous Methods for Deep Reinforcement Learning. In *Proceedings of The 33rd International Conference on Machine Learning*, pp. 1928–1937. PMLR, June 2016.

Nassar, M. R., Wilson, R. C., Heasly, B., and Gold, J. I. An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, 30(37):12366–12378, 2010.

Nickerson, R. S. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220, 1998.

Ortega, P. A., Wang, J. X., Rowland, M., Genewein, T., Kurth-Nelson, Z., Pascanu, R., Heess, N., Veness, J., Pritzel, A., Sprechmann, P., et al. Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030*, 2019.

Palminteri, S. and Lebreton, M. The computational roots of positivity and confirmation biases in reinforcement learning. *Trends in Cognitive Sciences*, 26(7):607–621, July 2022. ISSN 13646613. doi: 10.1016/j.tics.2022.04.005.

Palminteri, S., Lefebvre, G., Kilford, E. J., and Blakemore, S.-J. Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLOS Computational Biology*, 13(8):e1005684, August 2017. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005684.

Pedersen, M. L., Frank, M. J., and Biele, G. The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic bulletin & review*, 24:1234–1251, 2017.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Rescorla, R. and Wagner, A. A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. *Classical Conditioning: Current Research and Theory*, January 1972.

Roscher, R., Bohn, B., Duarte, M. F., and Garcke, J. Explainable machine learning for scientific insights and discoveries. *Ieee Access*, 8:42200–42216, 2020.

Roziere, B., Gehring, J., Gloeckle, F., Sootla, S., Gat, I., Tan, X. E., Adi, Y., Liu, J., Remez, T., Rapin, J., et al. Code llama: Open foundation models for code. *arXiv preprint arXiv:2308.12950*, 2023.

Schulz, E., Bhui, R., Love, B. C., Brier, B., Todd, M. T., and Gershman, S. J. Structured, uncertainty-driven exploration in real-world consumer choice. *Proceedings of the National Academy of Sciences*, 116(28):13903–13908, July 2019. doi: 10.1073/pnas.1821028116.

Schwarz, G. Estimating the dimension of a model. *The annals of statistics*, pp. 461–464, 1978.

Sharot, T. The optimism bias. *Current Biology*, 21(23): R941–R945, December 2011. ISSN 0960-9822. doi: 10.1016/j.cub.2011.10.030.

Shepperd, J. A., Klein, W. M. P., Waters, E. A., and Weinstein, N. D. Taking Stock of Unrealistic Optimism. *Perspectives on Psychological Science*, 8(4): 395–411, July 2013. ISSN 1745-6916. doi: 10.1177/1745691613485247.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Von Oswald, J., Niklasson, E., Randazzo, E., Sacramento, J., Mordvintsev, A., Zhmoginov, A., and Vladymyrov, M. Transformers learn in-context by gradient descent. In *International Conference on Machine Learning*, pp. 35151–35174. PMLR, 2023.

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., and Anandkumar, A. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*, 2023.

Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., Blundell, C., Kumaran, D., and Botvinick, M. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*, 2016.

Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., Blundell, C., Kumaran, D., and Botvinick, M. Learning to reinforcement learn, January 2017.

Wei, J., Wei, J., Tay, Y., Tran, D., Webson, A., Lu, Y., Chen, X., Liu, H., Huang, D., Zhou, D., et al. Larger language models do in-context learning differently. *arXiv preprint arXiv:2303.03846*, 2023.

Williams, R. and Peng, J. Function Optimization Using Connectionist Reinforcement Learning Algorithms. *Connection Science*, 3:241, September 1991. doi: 10.1080/09540099108946587.

Wilson, R. C. and Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *eLife*, 8: e49547, November 2019. ISSN 2050-084X. doi: 10.7554/eLife.49547.

Wu, H., Fai Cheung, S., and On Leung, S. Simple use of bic to assess model selection uncertainty: An illustration using mediation and moderation models. *Multivariate behavioral research*, 55(1):1–16, 2020.

Xie, S. M., Raghunathan, A., Liang, P., and Ma, T. An explanation of in-context learning as implicit bayesian inference. *arXiv preprint arXiv:2111.02080*, 2021.

Zhang, S. and Yu, A. J. Forgetful bayes and myopic planning: Human learning and decision-making in a bandit setting. *Advances in neural information processing systems*, 26, 2013.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.

A. Methods: Additional details

A.1. Prompt engineering

We point out the two design decisions that were crucial to achieving an above-chance-level performance on all tasks: (1) We framed this task in a gambling context as this allowed us to explicitly mention the rewards the agent can expect and that they vary stochastically with different probabilities. (2) We randomly sample the letters for each slot machine for each task. We excluded certain letters (I, U, X, Y, Z) as we noticed that the LLM was positively (I, U) or negatively (X, Y, Z) biased towards choosing them.

A.2. Parameter estimation

The model parameters were fitted using a maximum a posteriori approach. For the RW model, this involves fitting parameters $\theta = [\alpha, \beta]$ (two parameters in total). For the RW$\pm$ model, this involves fitting parameters $\theta = [\alpha^+, \alpha^-, \beta]$ (three parameters in total). The model parameters are fitted separately for each simulated run using the following objective:

$$\hat{\theta} = \arg\min_{\theta} \left[-\ln p(\theta) - \sum_{n=1}^{N} \ln p(a_n|\theta) \right]$$

Prior probabilities were based on (Daw et al., 2011), using a beta distribution $\mathcal{B}(1.1, 1.1)$ for learning rates and a gamma distribution with $\mathcal{G}(1.2, 5)$ for the inverse temperature parameter. We used a bound-constrained minimization that was implemented using scipy's `minimize` function, which internally relies on the L-BFGS-B algorithm (Zhu et al., 1997).

Parameter values were initialized in the range of 0 to 1 for α^+ and α^- and 0 to 10 for β. The fitting procedure was repeated 100 times and the time it took varied depending on the number of free parameters of the cognitive model. For one simulated run, the procedure took between ten seconds and three minutes on a standard desktop computer.

B. Generalizability across different LLMs and task formats

B.1. Further LLMs

We performed the partial feedback task on seven additional LLMs. Namely, we tested Claude-2.1, Claude-3 Haiku, GPT-4, and four versions of Llama-2, i.e. the 7 billion and the 70 billion parameter models with and without RLHF (Llama-2-7B, Llama-2-7B-Chat, Llama-2-70B, and Llama-2-70B-Chat). We analyzed the learning behavior of these models and found that the observed learning asymmetry is not unique to Claude-1.2 but is also significant in all tested LLMs (see Figure 6).

We also tested the influence of RLHF and model size on the emergence of the optimism bias. We used an ordinary least squares linear regression model for the Llama-2 model family, since the exact sizes for GPT-4 and the Claude models are unknown. We defined the optimism bias using the difference in learning rates (i.e., $\alpha^+ - \alpha^-$) as our dependent variable. We included RLHF, coded as 1 for presence and 0 for absence, model size (7 or 70), and a constant bias term as independent variables in the regression model. Our results show that RLHF tends to increase the optimism bias (with a coefficient of 0.08, $p < 0.001$), while an increase in model size tends to decrease this effect (with a coefficient of -0.02, $p < 0.001$). However, it is important to note that the scope of this analysis was limited to the Llama-2 models, which limits our ability to generalize these results to other models, such as GPT-4 or Claude.

B.2. Further task formats

We tested Claude-1.2 on variations of the partial feedback task to check the robustness of our results. We manipulated two aspects: For one, we added one or two additional slot machines to each casino, leading to three- and four-armed bandit problems. Furthermore, we used two different reward magnitudes for success and failure (0.5 and -0.5 and 1.0 and 0.0). This resulted in six new task formats. We found that the optimism bias remains persistent across these variations (see Figure7). This demonstrates that the underlying optimistic learning dynamics generalize to settings with different reward magnitudes and more than two options.

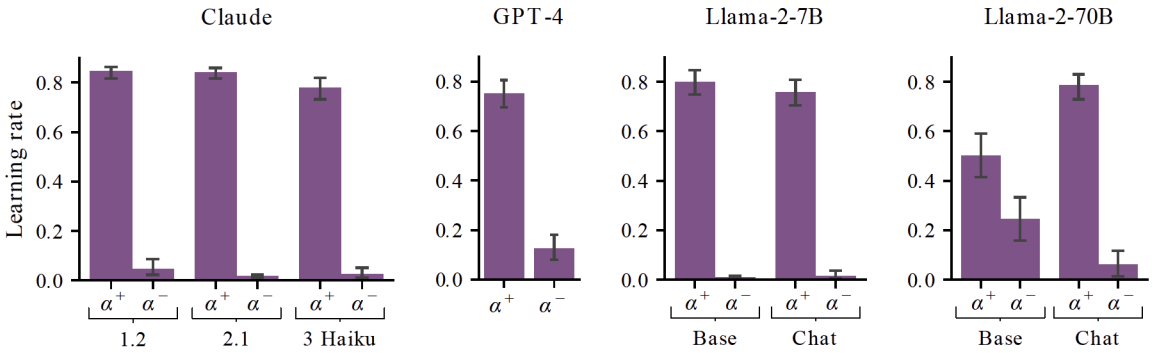

Figure 6. Learning rate comparison of eight LLMs in the partial feedback task. All analyzed LLMs show a significant learning asymmetry with a stronger response to positive prediction errors than to negative prediction errors, indicating an optimism bias. Plots are divided by model family. Error bars correspond to 95% CIs.

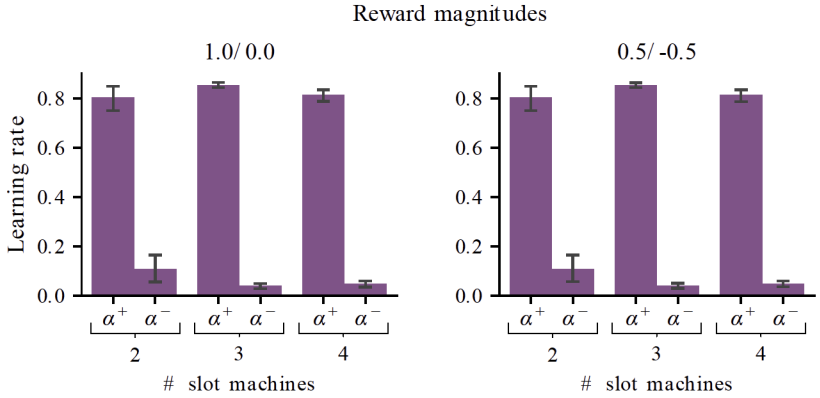

Figure 7. LLM learning rate analysis for partial feedback task variations. Slot machines returned two different pairs of rewards for success and failure of either 1.0 and 0.0 (left) or 0.5 and -0.5 (right). In addition, the casinos contained a varying number of slot machines (2, 3, and 4). The LLM shows an optimism bias in all settings tested. Error bars correspond to 95% CIs.

C. Task 2: Additional details

C.1. Prompts

The full information task consisted of 16 casinos, half of which were partial feedback casinos and half of which were full feedback casinos. Below is a sample prompt for both types of casinos. The differences between the partial and full feedback casinos are shown in bold.

Prompt Task 2: partial feedback casinos

You will visit a casino 40 times. The casino has two slot machines that stochastically return either 1 or -1 with different reward probabilities. You can only interact with one slot machine per visit. Half of the time you visit the casino, you can play, the other someone else is playing. During visits where you can play, you'll earn points from the chosen machine. During visits where someone else is playing, you'll learn what points are earned on the chosen machine. Your goal is to maximize the total amount of points you receive in all 20 visits you can play.

During your previous visits you have observed the following:

- On visit 1 someone else played Machine H and earned 1.0 point.
- On visit 2 you played Machine H and earned 1.0 point.
- On visit 3 you played Machine E and earned -1.0 point.

Q: You are now in visit 4. Which machine do you choose between Machine E and Machine H?
A: Machine [insert]

Prompt Task 2: full feedback casinos

You will visit a casino 40 times. The casino has two slot machines that stochastically return either 1 or -1 with different reward probabilities. You can only interact with one slot machine per visit. Half of the time you visit the casino, you can play, the other someone else is playing. During visits where you, can play, you'll earn points from the chosen machine. **You'll also learn what points would have been earned had the other machine been selected.** During visits where someone else is playing, you'll learn what points are earned on the chosen **and what points would have been earned had the other machine been selected. Nevertheless, you only accumulate points from the machine you choose to play.** Your goal is to maximize the total amount of points you receive in all 20 visits you can play.

During your previous visits you have observed the following:

- On visit 1 someone else played Machine H and earned 1.0 point.
  On Machine E the player would have earned -1.0 point.
- On visit 2 you played Machine H and earned 1.0 point.
  On Machine E you would have earned -1.0 point.
- On visit 3 you played Machine E and earned -1.0 point.
  On Machine H you would have earned 1.0 point.

Q: You are now in visit 4. Which machine do you choose between Machine E and Machine H?
A: Machine [insert]

C.2. Model comparison

We simplified the analysis of the original by fitting separate learning rates only for the chosen and unchosen option, but not separate learning rates for free and forced-choice trials. Furthermore, we used only the simulated behavior from full feedback casinos for the fitting of two cognitive models – a 2α and a 4α model.

The model comparison revealed that the 2α model provided a better fit to the behavior. The 2α model contained only two learning rates – combining the learning rates for the chosen and unchosen options, which either confirmed or disconfirmed prior beliefs. As can be seen in Figure 8, all agents show a clear tendency to overweight information that confirms their choices.

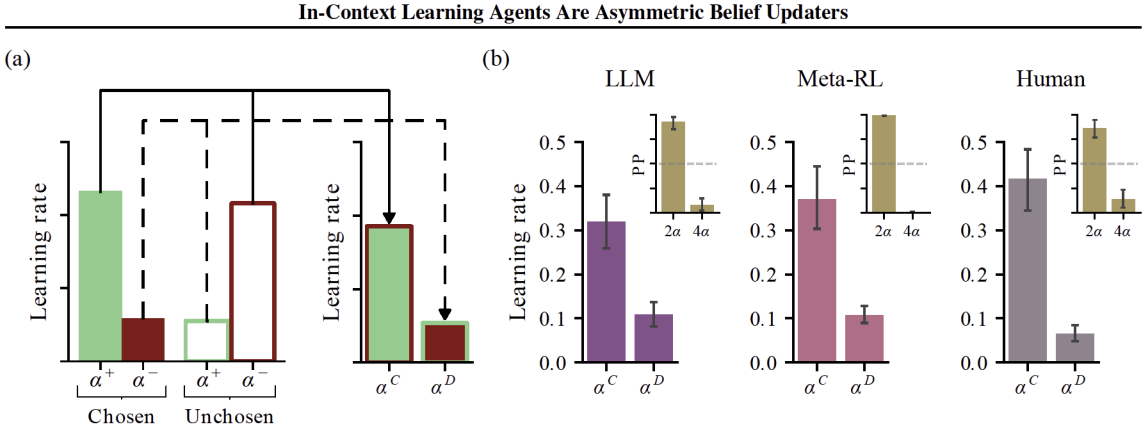

Figure 8. Confirmation bias in the full feedback task. (a) Schematic showing how the learning rates of the full model (i.e. a model for a different learning rate for each possible combination of outcome types and prediction error types) relate to those of the confirmation bias model (2α), which bundles together the learning rates for positive chosen and negative unchosen (i.e. confirmatory) prediction errors (α^C) and the learning rates for negative chosen and positive unchosen (i.e. disconfirmatory) prediction errors (α^D). Adapted from (Palminteri & Lebreton, 2022) (b) Fitted confirmation bias model for the LLM, the Meta-RL agent and human participants. The average posterior probabilities (PP) indicate that the 2α model is a superior fit in all model comparisons. Error bars correspond to 95% CIs.

D. Task 3: Additional details

D.1. Prompts

The agency condition consisted of 12 casinos and provided only partial feedback. Half of these casinos were free-choice casinos containing only free-choice trials, and the other half were a mixture of free-choice and forced-choice trials. The free-choice casinos consisted of 40 trials per casino. The mixed-choice casinos consisted of 80 trials with 40 free-choice trials and 40 forced-choice trials. Below is a sample prompt for both types of casinos. The differences between the free-choice and mixed-choice casinos are shown in bold.

Prompt Task 3: free-choice casinos

You will visit a casino 40 times. The casino has two slot machines that stochastically return either 1 or -1 with different reward probabilities. You can only interact with one slot machine per visit. Your goal is to maximize the total amount of points you receive in all 40 visits you can play.

During your previous visits you have observed the following:

- On visit 1 you played Machine H and earned 1.0 point.
- On visit 2 you played Machine N and earned -1.0 point.
- On visit 3 you played Machine H and earned -1.0 point.

Q: You are now in visit 4. Which machine do you choose between Machine N and Machine H?
A: Machine [insert]

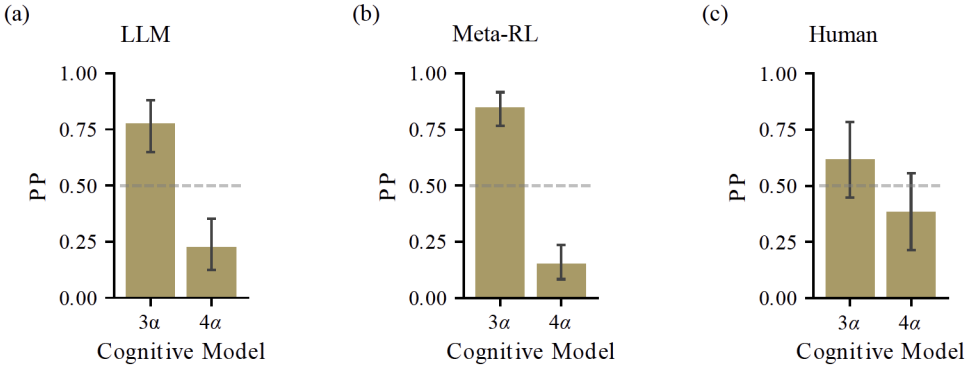

Figure 9. Model comparison for the agency condition for all agents. The 3α model has the highest average posterior probabilities (PP). This indicates that information from forced-choice trials where no agency is involved are weighted symmetrically whereas the free-choices are weighted asymmetrically. Error bars correspond to 95% CIs.

Prompt Task 3: mixed-choice casinos

You will visit a casino 80 times. The casino has two slot machines that stochastically return either 1 or -1 with different reward probabilities. You can only interact with one slot machine per visit. **Half of the time you visit the casino, you can play, the other half someone else is playing and you can only see the rewards for their chosen slot machine.** Your goal is to maximize the total amount of points you receive in all 40 visits you can play.

During your previous visits you have observed the following:

- On visit 1 you played Machine H and earned 1.0 point.
- **On visit 2 someone else played Machine N and received -1.0 point.**
- On visit 3 you played Machine H and earned -1.0 point.

Q: You are now in visit 4. Which machine do you choose between Machine N and Machine H?
A: Machine [insert]

D.2. Model comparison

We only fit two cognitive models to the mixed-choice casinos. One model consisted of separate learning rates for positive and negative prediction errors for free- and forced-choice trials (4α). The second model consisted of two learning rates for free-choice trials and only one learning rate for forced-choice trials (3α). Model parameters were fit based on the free-choice trials of the mixed-choice casinos. Model comparison indicated that for all agents, the 3α model provided a better fit to the data based on the average PP (see Figure 9).

E. Meta-RL: Additional details

E.1. Agent

The agent consisted of a Transformer network with a model dimension of $8 \cdot \text{input\_size}$, two feedforward layers with a dimension of 128, and eight attention heads, followed by two linear layers that output a policy and a value estimate, respectively. The agent received the previously selected action (one-hot encoded) and the reward of that action, alongside the current context. This context included a normalized time index for all tasks. For Task 1, a one-hot encoded representation of the four casinos was added, resulting in an input size of eight. The context of Tasks 2 and 3 included a bit representation of the trial type (i.e. 00 = free-choice, 10 = forced-choice left option chosen, 01 = forced-choice right option chosen) for the current and previous trial. Task 2 additionally included the reward of the current unchosen option, resulting in an input size of eight for Task 2 and nine for Task 3. In the partial feedback casinos of Task 2, a placeholder of 0 for the missing reward signal of the unchosen option was propagated. To prevent learning from forced-choice trials, we masked the policy and

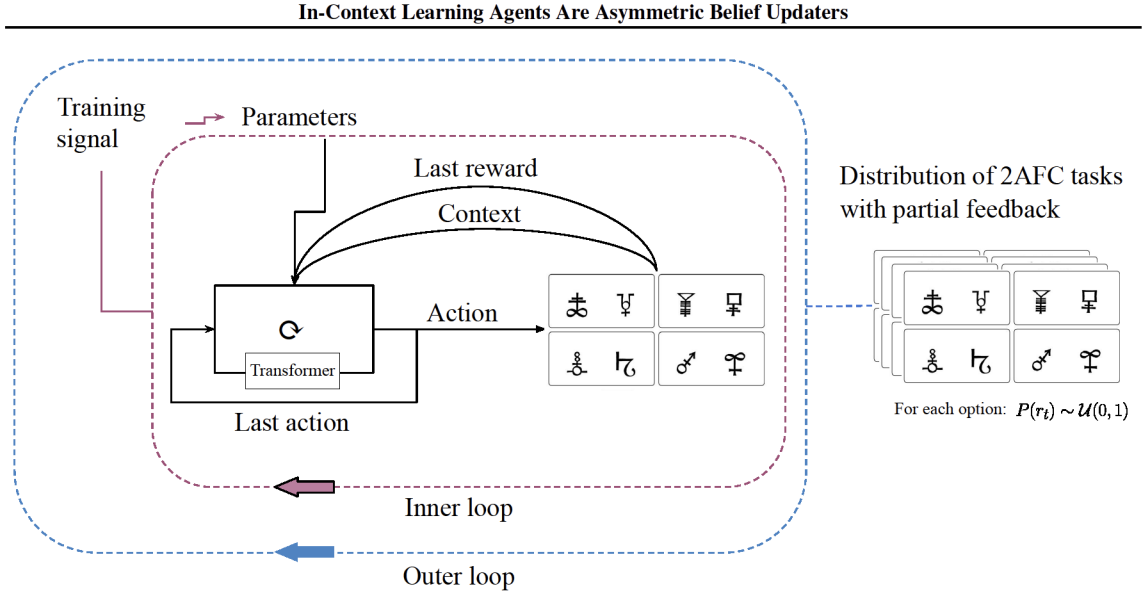

Figure 10. Schematic of the Meta-RL training for Task 1 highlighting the inner and outer training loops. The outer loop adjusts the weights of the Transformer in response to the learning experience. These weights shape the behavior of the Transformer in the inner loop as it interacts with 2AFC tasks (here Task 1). After each cycle of the outer loop, a new task is sampled where the probabilities for each option are sampled from a uniform distribution. Adapted from Botvinick et al. (2019).

value loss for those trials. To ensure a consistent input dimension, we use placeholder values for the initial inputs and all the unseen trials.

Network weights were adjusted by gradient-based optimization using ADAM (Kingma & Ba, 2014) on a standard actor-critic loss (Mnih et al., 2016) at the end of each training episode. The initial learning rate for ADAM was 0.0003. For the actor-critic loss, we used a discount factor of 0.8 and weighted the critic loss with 0.5. In addition, we used entropy regularization to discourage premature convergence to sub-optimal policies (Williams & Peng, 1991). Starting with an entropy coefficient of 1, we linearly decayed the influence of the entropy term to 0 after half of the $5,000$ episodes. We used a batch size of 64 during training.

E.2. Training

The training process of the Meta-RL agent is graphically depicted in Figure 10. In the process, there are two optimization loops – an outer and an inner loop. At the beginning of each training episode (outer loop), we sample a new task b_i from the prior distribution of 2AFC tasks. As stated earlier, we assumed a uniform distribution for the win probabilities for each option in each task. The goal of the agent is to find a history-dependent policy $\pi_{\boldsymbol{\Theta}}$ that maximizes the expected total discounted reward accumulated during an episode. For that, the parameters $\boldsymbol{\Theta}$ are adjusted at the end of each episode. Since the decision strategy changes across training episodes, the agent must act differently according to its prior belief over which part of the task distribution it is currently in. As the optimization maximizes the expected total rewards across tasks, the policy starts to generalize the underlying principles that help reach this objective.

The agent interacts in the inner loop with the specific sampled task b_i, aiming to maximize its rewards across all steps with the help of its policy. At the beginning of each step, a context c_t is drawn from a uniform distribution. Upon receiving an action a_t, the environment computes a reward r_t and samples the next context c_{t+1} to which the agent steps forward. The next context c_{t+1}, the action a_t, and the reward r_t are concatenated and added to the input to the Transformer. As demonstrated by Wang et al. (2017), this input design is crucial for an agent to learn an association between choices that have been made in particular states and their subsequent rewards.

After training to convergence, we tested the Meta-RL agent on the three experimental task under the same conditions as with the LLM. This idealized in-context learning agent outperformed the human as well as the LLM (see Figure 11).

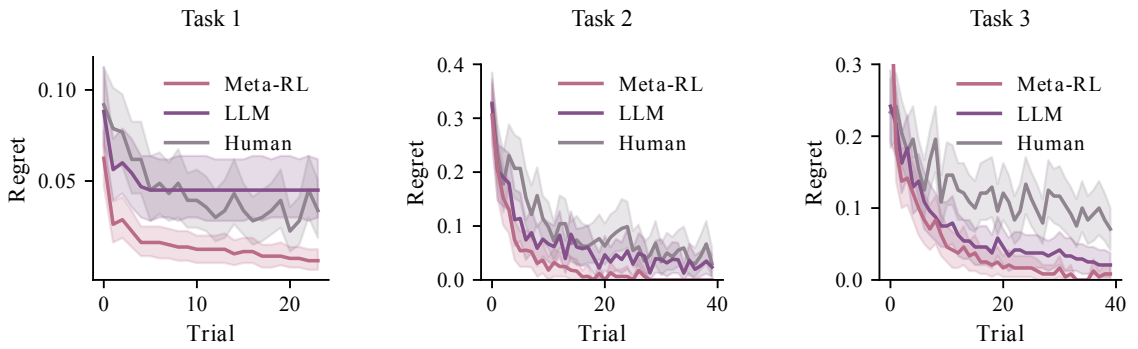

Figure 11. Average performance comparison of all three agents in terms of regret. In all tasks, the Meta-RL agent improves its performance over time fastest. Shaded areas correspond to 95% CIs.

F. Data availability

Our code is publicly available at `https://github.com/jschbrt/InContext-Learning-Dynamics`.

2.3 A resource-rational account of zero-shot compositional inference in a reinforcement learning setting

Jagadish, A. K., Binz, M., Saanum, T., Wang, J. X., Schulz, E. (2024). A resource-rational account of zero-shot compositional inference in a reinforcement learning setting. psyarxiv:ymve5.

Contributions in-context

Humans have a remarkable ability to generalize from limited experiences, combining prior knowledge in novel ways to instantly tackle new situations. In this study, we investigated the human propensity to reason compositionally in a reinforcement learning setting, with a particular focus on improving our understanding of how people make such zero-shot compositional inferences.

We first developed a new experimental paradigm in which agents, natural or artificial, must learn the reward functions underlying different options of the bandit task in two sub-tasks and later combine them to solve a third sub-task. We considered two reward composition rules: (1) additive, where rewards are summed over sub-tasks; (2) change point, where the reward functions switch from one function to the other after a specific option. Then, we derived a meta-reinforcement learning agent (RL^2) [47, 128] capable of performing optimal zero-shot compositional inference. We demonstrated this optimality by comparing it with the compositional Gaussian process regression model, which served as the true generative model of the task. This finding illustrated that meta-learning facilitates compositional reasoning in neural networks; see Section 3.1 in Outlook for an extended discussion.

We subsequently conducted experiments on human participants and found that they can indeed make such zero-shot inferences, but they do not always do so optimally. To account for these deviations from optimality, we introduced a resource-rational extension of the meta-reinforcement learning agent (RR-RL^2) [129] (Section 1.6 in Background for an overview of resource-rational analysis). The central hypothesis was that participants strive for optimal performance but are limited by cognitive constraints, leading them to rely on algorithms with bounded computational complexity. The resource-rational agent is perfectly placed to model such behavior. For comparison, we evaluated six other Bayesian models that vary along three dimensions: (1) whether they can generalize learned values from one option to the other; (2) whether they use an uncertainty-guided exploration strategy; (3) whether they can compose the learned values from the first two sub-task to solve the final sub-task. Model comparison revealed that RR-RL^2 best captured human compositional inference. Specifically, we found that the parameter controlling for computational complexity, when fitted to participants, provided a reliable measure of participants' capacity for zero-shot compositional reasoning.

Finally, we discussed how these findings shed light on performing compositional reasoning with neural networks (see Section 3.2 in Outlook for

an extended discussion), their relation to other studies in compositional reinforcement learning, the role limited resources play in compositional reasoning, and possible neural and cognitive processes underlying it. We concluded by acknowledging the limitations of experimental design, and by highlighting how these results contribute to a broader effort to study compositional reasoning in humans, symbolic systems, and sub-symbolic systems under computational constraints.

A resource-rational account of zero-shot compositional inference in a reinforcement learning setting

Akshay K. Jagadish[1,2,+], Marcel Binz[1,2], Tankred Saanum[1], Jane X. Wang[3], and Eric Schulz[1,2,*]

[1]MPRG Computational Principles of Intelligence, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2]Institute for Human-Centered AI, Helmholtz Computational Health Center, Munich, Germany
[3]Google Deepmind, London, United Kingdom
[+]akshay.jagadish@tue.mpg.de
[*]eric.schulz@helmholtz-munich.de

ABSTRACT

People can easily evoke previously learned concepts, compose them, and apply the result to solve novel tasks on the first attempt. The aim of this paper is to improve our understanding of how people make such zero-shot compositional inferences in a reinforcement learning setting. To achieve this, we introduce an experimental paradigm where people learn two latent reward functions and need to compose them correctly to solve a novel task. We find that people have the capability to engage in zero-shot compositional reinforcement learning but deviate systematically from optimality. However, their mistakes are structured and can be explained by their performance in the sub-tasks leading up to the composition. Through extensive model-based analyses, we found that a meta-learned neural network model that accounts for limited computational resources best captures participants' behaviour. Moreover, the amount of computational resources this model identified reliably quantifies how good individual participants are at zero-shot compositional reasoning. Taken together, our work takes a considerable step towards studying compositional reasoning in agents – both natural and artificial – with limited computational resources.

Introduction

People have an impressive ability to learn from sparse data[1]. We can acquire a new word from only one encounter[2]. We can achieve near-perfect classification rates from only one labelled observation[3]. We can even ask questions such as "how likely is it that a newly invented machine could transform a man into a vase?"[4], even though we are unlikely to ever encounter such a machine. Many researchers have proposed that the ability to generalize from sparse data is a hallmark of human intelligence[5,6].

What are the mechanisms that underlie this ability? One mechanism that enables strong generalizations is compositionality[5,7–10], which is the idea that complex entities can be constructed through the combination of primitive elements. People are generally considered to excel at reasoning compositionally[11]. They can, for example, combine parts of objects into novel objects[7,12] or compose previously learned actions to explore in novel contexts[8,13]. It has thus been argued that compositionality equips us with the ability to *"make infinite use of finite means"*[14,15], allowing us to generalize to novel situations by reusing and combining past experiences[13,16,17].

Empirical studies have demonstrated that people have an inherent predisposition towards compositional patterns[7,18–23]. For example, utilizing the function learning paradigm, which involves the learning, completion, and prediction of functional patterns, Schulz and colleagues[18,19] have demonstrated that humans find it easier to learn about compositional than non-compositional patterns. Furthermore, they showed that humans exhibit superior abilities to complete and predict compositional functions, as well as an enhanced capacity for remembering such functions[20,21]. These findings extend beyond function learning to other domains such as spatial structure learning[23,24], concept learning[7,25], shape perception[26], and auditory sequence learning[27]. Taken together, there is strong evidence for the presence of compositional inductive biases in humans.

While the preference for compositional patterns has received significant attention, how people compose two already learned functions and act on them in a zero-shot manner remains less well-understood. We attempt to close this gap in the present paper by studying human compositional reasoning in a reinforcement learning setting. More specifically, we are interested in how people apply a compositional rule instructed to them on learned latent reward functions. To study this question, we propose a novel experimental paradigm in which people interact with a sequence of three structured multi-armed bandit tasks[20,28–31] as

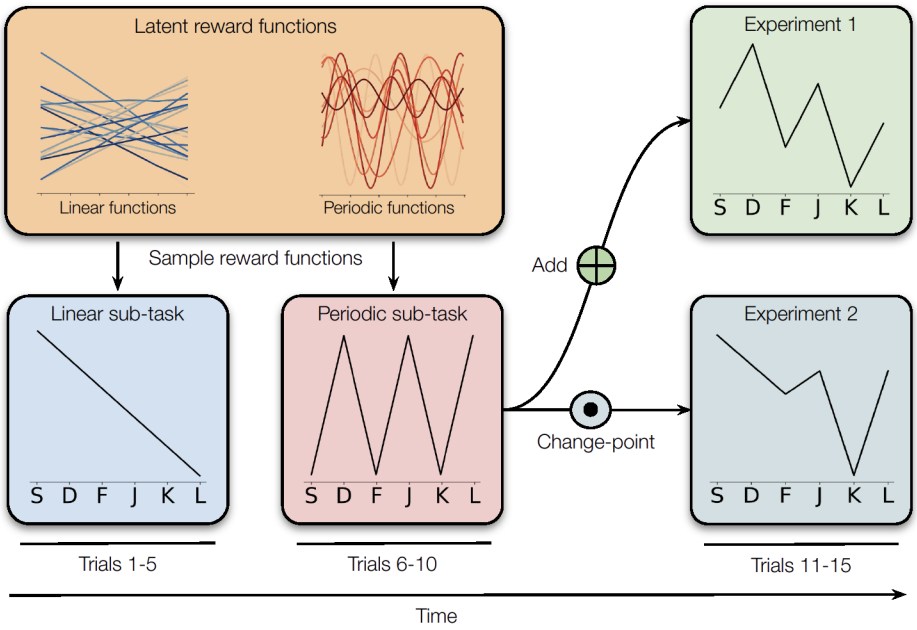

Figure 1. Overview of our experimental paradigm. Each task consists of three multi-armed bandit sub-tasks, with participants performing five trials per sub-task. Each multi-armed bandit task has six arms, each corresponding to different letters on the keyboard (starting from S to L). Rewards for the arms in our bandit tasks follow a latent function that is dependent on the spatial position of the arms. The reward functions for the first two sub-tasks are sampled from either the linear or the periodic family. The rewards for the final sub-task are constructed by composing the reward functions sampled in the two earlier sub-tasks. We conducted experiments with two different composition rules: an additive rule (experiment 1) and a change-point rule (experiment 2). In experiment 1, rewards for the final sub-task are constructed by performing an element-wise summation of sampled rewards from the first two sub-tasks. Whereas in experiment 2, they are constructed by combining segments of the sampled rewards such that the rewards change from being from one family to another after a certain number of options. Specifically, rewards in the first segment (i.e. options S-D-F) come from one family – either a linear or periodic family – and the alternative family in the second segment (i.e. options J-K-L). Note that the order in which these reward segments are combined in the change-point rule is randomized. If participants learn the latent reward functions in the first two sub-tasks and apply the compositional rule correctly, they can – in principle – perform zero-shot compositional inference, meaning that they choose the optimal arm (arm D in the example shown for experiment 1 and arm S in the example shown for experiment 2) on the first trial of the last sub-task.

illustrated in Fig. 1. The rewards for the first two sub-tasks are sampled from differently structured functions. They are followed by a third sub-task in which the rewards are set to a composition of the previously encountered functions. The structure of our task induces a learning curriculum that allows participants to rapidly solve the final sub-task in a zero-shot manner – assuming that they apply the instructed compositional rule.

In two experimental studies, we find that people can learn the underlying latent functions via reinforcement learning and apply the compositional rule over the learned functions in zero-shot. They do so rapidly right from the first task onward, a propensity studied under the paradigm of rapid instructed task learning[32,33]. However, our analyses also indicate that their behaviour deviates systematically from a fully-normative account. Extensive model-based analyses furthermore reveal that human compositional reasoning is overall best explained by a resource-rational account[34,35]. Taken together, our results suggest that people can make zero-shot compositional inferences but that their performance is constrained by cognitive demands.

To investigate compositional reinforcement learning in humans, we developed a novel multi-armed bandit paradigm based on previous works[20,29,30] as illustrated in Fig. 1. Each task consists of three multi-armed bandit sub-tasks in which rewards follow a latent function that is dependent on the spatial position of the arms. The reward functions for the first two sub-tasks are sampled from either the linear or the periodic family of functions. In the final sub-task, reward functions are constructed by composing the reward functions encountered in the two earlier sub-tasks. We conducted two experiments with different

composition rules: an additive rule and a change-point rule. Our task induces a learning curriculum, which enables us to probe whether people are able to reason compositionally in a reinforcement learning setting. In particular, we expect people to solve the final sub-task in a zero-shot manner, selecting the best option on the first trial. To have a comparison, we also consider a condition without a curriculum. In this non-curriculum condition, people do not interact with the first two sub-tasks and instead directly observe the composite function from the final sub-task. We set the length of each sub-task to five, leading to 15 trials per task in the curriculum and five trials per task in the non-curriculum condition. Note that the number of trials per sub-task is less than the number of available options, thereby preventing participants from exhaustively trying out all options and forcing them to generalize based on the underlying function.

Experiment 1: Additive rule

We conducted an online behavioural study, on the Prolific platform, following the structure of the just outlined task to test the underlying mechanisms behind how people compose. Participants played a game under a cover story that they were interacting with slot machines produced by two manufacturers (Blue Lagoon and Green Geeks). They were told that all slot machines from the same manufacturer behaved similarly. However, participants were not told which manufacturer each slot machine belonged to, but had to figure this out through trial and error. In the curriculum condition, participants played with a slot machine from each manufacturer before playing a compositional slot machine that combined the two. In the non-curriculum condition, participants only played with the compositional slot machine. The study involved 20 tasks per participant, leading to 300 trials in total for the curriculum condition and 100 for the non-curriculum condition. We provide further details about the experiment and participants in the Materials and Methods section.

Behavioural analysis

First, we wanted to establish that people successfully composed reward functions in our task. For the corresponding analyses, we considered two behavioural measures: regrets and the probability of making optimal choices. Fig. 2(a) shows the mean regret of participants in the compositional sub-task. The regret is computed by taking the difference between the highest reward in the given sub-task and the reward for the action selected by the participant. Participants are said to have successfully performed zero-shot compositional inference if they choose the optimal arm on the first trial of the compositional sub-task (or have a regret measure of 0 on the first trial). We see from the regrets that people performed better than chance right from the outset for the curriculum condition (Mean $(M) = 2.163$, Standard Error $(SE) = 0.116$; t[1] $= -19.57, p < 0.001$) whereas they start at chance-level for the non-curriculum condition ($M = 4.055$, $SE = 0.059$). To further quantify the effects of zero-shot compositional inference, we performed a mixed-effects linear regression on the regrets in the final sub-task with trials, conditions, and their interaction as fixed effects (with random slopes and intercepts per participant for all of these factors). This analysis revealed that participants in the curriculum condition had a significantly lower regret on the first trial of the final sub-task than participants in the non-curriculum condition ($\hat{\beta} = -1.18 \pm 0.115; z = -10.24, p < 0.001$). In addition, a comparison of the probability for making an optimal choice on the first trial between curriculum and non-curriculum conditions, shown in Fig. 2(b), confirmed that people in the curriculum condition ($M = 0.382$, $SE = 0.02$) made optimal choices more frequently than in the non-curriculum condition ($M = 0.192$, $SE = 0.07$; $t = -9.242, p < 0.001$). Regret performance on the first trial of the curriculum condition was even better than performance on the last trial of the non-curriculum condition ($M = 2.40$, $SE = 0.13$; $t = 2.72, p < 0.01$), suggesting that learning within the last sub-task cannot match the performance boost gained from compositional inference.

While people were able to compose in a zero-shot manner, they did not do so perfectly. Their initial regrets in the final sub-task ($M = 2.163$, $SE = 0.116$; $t = 34.60, p < 0.001$) deviated significantly from ideal compositional reasoning. Further evidence of people's suboptimality comes from the observation that they continued learning during the final sub-task in the curriculum condition (which would not be needed if they were to engage in perfect zero-shot compositional inference). To quantify this effect, we fitted a mixed-effects linear regression model using per-trial regret in the last sub-tasks as the dependent variable, and the corresponding trial number as both fixed effects and random effects over participants. The results of this model showed a significant fixed effect of trial number ($\hat{\beta} = -0.32 \pm 0.02; z = -13.88, p < 0.001$) onto regret, confirming that the performance of participants improved with additional interactions. The observed improvement in the curriculum condition ($\hat{\beta} = -0.21 \pm 0.02; z = -12.26, p < 0.001$) was generally weaker than that in the non-curriculum condition ($\hat{\beta} = -0.42 \pm 0.02; z = -21.80, p < 0.001$).

We also inspected the marginal action distribution of participants on the first trial of the final sub-task shown in Fig. 2(c). We see that the mode of the participants' action distribution matches the optimal choice, but that human behaviour also systematically deviates from optimal behaviour. Particularly, one interesting feature is that people seem to pick corner arms – especially the left-most one – frequently. This could reflect a bias that has been observed in other studies of people exploring different options starting from left to right[29]. However, probability for making an optimal choice on the first trial was still

[1] t-values are reported from a non-parametric independent two-sample t-test (two-tailed) using 1000 random permutations.

better in curriculum condition than in the non-curriculum condition even when only compositions where non-corner arms were optimal was considered (for further details, see SI). This indicates that despite their bias towards choosing the corner arms, participants are still able to perform zero-shot compositional inference.

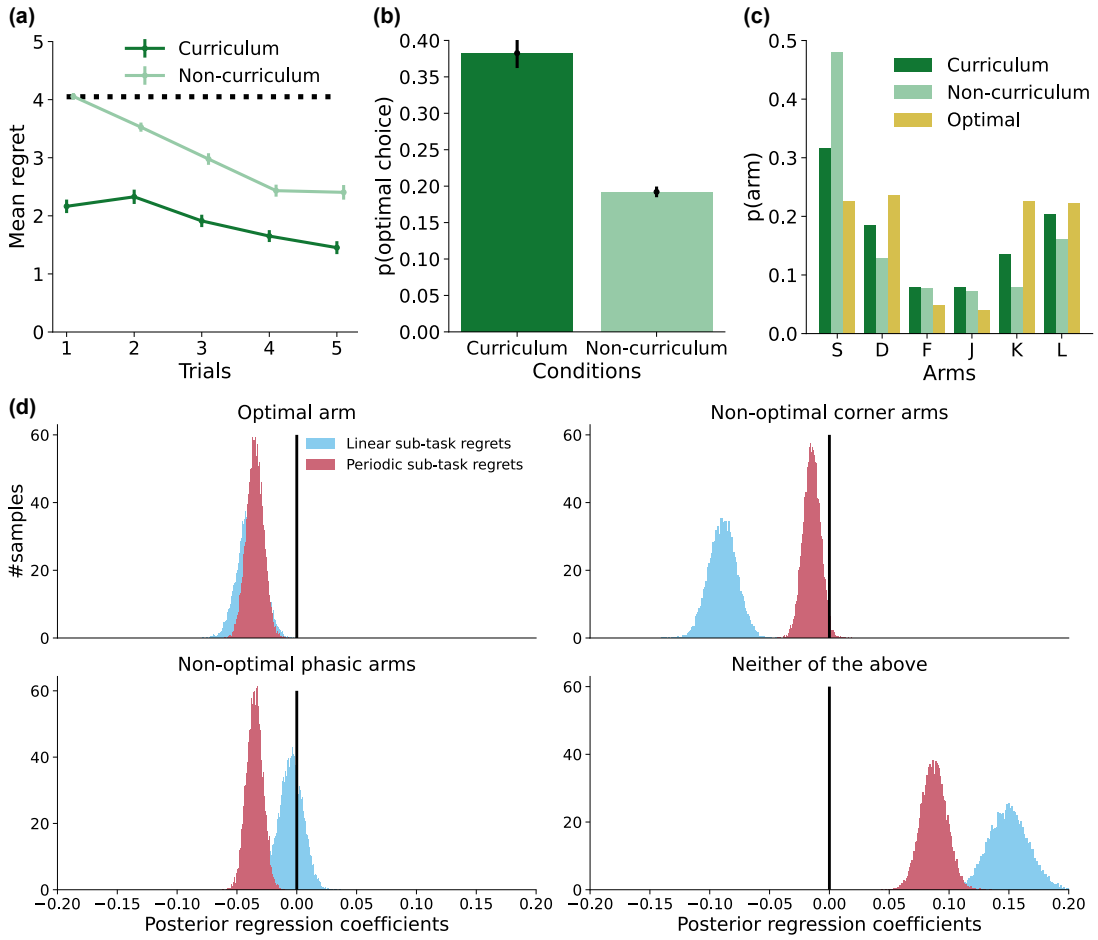

Figure 2. Behavioural results of experiment 1. (a) Mean regrets for participants in the final sub-task for the two conditions: curriculum and non-curriculum. The dotted lines in black indicate mean regret from a random policy. (b) Probability of participants making the optimal choice on the first trial of the final sub-task for the curriculum and non-curriculum condition. Error bars in (a) and (b) represent standard errors computed over participants. (c) Marginal distribution of choices on the first trial of the final sub-task for the two conditions. The bar in gold shows the marginal distribution of choices for the optimal policy. (d) Explaining choices of participants in the last sub-task based on their task performance in the first two sub-tasks. We classified the choices on the first trial into four categories: optimal arm (top-left), non-optimal corner arms (top-right), non-optimal phasic arm (bottom-left), and neither of the above (bottom-right). Then, we fit a Bayesian logistic regression model from the total regrets (summed over all trials) in the first two sub-tasks onto each of these categories. The sub-plots show the histogram of the posterior regression coefficients of linear and periodic sub-tasks for all four choice categories.

To better understand the mistakes that people make during compositional reasoning, we looked at how participants' performance in the first two sub-tasks can explain their behaviour on the final compositional sub-task. We first classified choices on the first trial of the compositional sub-task into four categories: first, picking the optimal arm as predicted by compositional inference; second, a non-optimal corner arms category which includes trials where people picked the corner arms despite them not being the optimal choice; third, a non-optimal phasic arms category which includes trials where arms

belonging to the same phase as the periodic sub-task were picked even when it was not the optimal choice; and fourth, a category which includes all trials where choices did not fall into any of the three categories mentioned above. We then fitted a separate Bayesian logistic regression model in PyMC3 from the total regrets (summed over all trials) in the first two sub-tasks onto each of these four choice categories coded as a binary variable. Fig. 2(d) visualized the posterior regression coefficients from these fitted models with each category shown in a separate sub-plot. When the dependent variable was picking the optimal choice, the posterior regression coefficients were negative with similar means for both linear ($M = -0.038$, $SE = 7.319e-05$) and periodic ($M = -0.035$, $SE = 4.925e-05$; $t = -40.728, p < 0.001$) sub-tasks. This suggests that participants pick the optimal arm when they learn both the sub-tasks well. The coefficients for regrets of the linear sub-tasks were more negative ($M = -0.089$, $SE = 8.107e-05$) than that of periodic regrets ($M = -0.015, SE = 5.040e-05$; $t = -779.036, p < 0.001$) when the dependent variable was non-optimal corner arms. This result suggests that people tend to pick non-optimal corners arms when they learned linear functions better than periodic functions. When the dependent variable was non-optimal phasic arms, the posterior regression coefficients for regrets from the periodic sub-task were lower ($M = -0.0352, SE = 0.477e-04$) than that of the linear sub-task ($M = -0.0048$, $SE = 7.0762e-05$; $t = 356.072, p < 0.001$). This result indicates that people pick one of the phasic arms from the periodic sub-task on the first trial of the compositional sub-task when they perform better in the periodic sub-task. Lastly, the regression coefficients were positive for both the regrets from the linear sub-task ($M = 0.1498$, $SE = 1.167e-05$; $t = 450.07, p < 0.001$) and those from the periodic sub-task ($M = 0.0874$, $SE = 7.502e-05$) when the dependent variable was neither of the categories above. This result suggests that people pick neither the optimal, the corner nor the phasic arm when they have not learned the underlying functions in either of the two sub-tasks well.

Taken together, behavioural results from experiment 1 suggest that people can compose in a zero-shot fashion but are not perfect. However, their mistakes are highly structured and can be predicted based on how well they have learned the different components of the first two sub-tasks.

Model-based analysis

In our compositional bandit task, people can – in principle – perform near-perfect if they manage to compose. However, the behavioural analysis above revealed that people (despite generally managing to compose) systematically deviated from optimal behaviour. To get a better understanding of these deviations and the cognitive processes behind them, we investigated people's behaviour using computational models.

We considered eight different computational models for explaining participants' choices in our task as summarised in Table 1. The models considered span five key axes in terms of the strategies people might be using: (1) from learning values for options independently to generalisation across options, (2) from value-driven learning to uncertainty-guided exploration, (3) from learning sub-tasks independently to composing learned reward functions over sub-tasks (4) from hand-designed priors to learned priors and (5) from unbounded compositional reasoning to resource-constrained reasoning.

Six of these are Bayesian models that vary along three dimensions: first, whether or not they can generalize learned values from one option to the other, second, whether or not they use uncertainty-guided exploration strategy, and third, whether or not they can compose the learned values from the first two sub-task to reason on the final sub-task. The Bayesian models include a Bayesian mean-tracker (BMT)[36] which is a model that does not learn about the underlying functional structure but instead updates its beliefs about rewards for each option independently, as well as a model that learns functions by generalizing across options within a sub-task based on the idea of Gaussian Process regression (GPR)[18,37,38]. For each of these two models, we considered one variant that cannot compose and instead learns separate reward functions for each sub-task, another that does not perform uncertainty-guided exploration, and lastly, one that initializes its predictions in the final sub-task to the composition of the learned means from the first two sub-tasks.

In addition, we also considered two recurrent neural network models that were trained via meta-reinforcement learning[39,40]. Unlike the Bayesian models from above, these models learn inductive biases about latent reward functions via trial-and-error, without requiring an explicit specification of priors[41]. The first of these models is RL^2 – a model that is known to approximate the Bayes-optimal policy for the distribution of tasks it was trained on[42,43], which thereby allows us to test whether people compose optimally. The second is a resource-rational extension of RL^2 referred to as RR-RL^2[44]. The particular resource constraint considered by RR-RL^2 is the description length of the meta-learned recurrent neural network, which is defined as the number of bits required to store its parameters. RR-RL^2 captures the hypothesis that people attempt to achieve optimal performance but that they are subject to the constraint of relying on an algorithm with limited computational complexity. We fitted RR-RL^2's description length on a participant-by-participant basis reflecting the assumption that different participants use different amounts of computational resources.

We simulated all models on our compositional bandit task and measured their performance in the final sub-task. We found that performance on the first trial was near-optimal for the five models that can compose (the compositional BMT and GPR, as well as the two meta-learning agents), indicating that they can re-use the earlier learned functions to compose new functions in a zero-shot manner; for detailed visualizations, see Supplementary Information (SI).

To obtain a quantitative measure of the goodness-of-fit of models to human choices, we conducted a Bayesian model

Table 1. The five key axes spanned by computational models of zero-shot compositional inference

Model	Generalization	Exploration	Composition	Learned priors	Bounded
Bayesian Mean-Tracker (mean only)	✗	✗	✗	✗	✗
Bayesian Mean-Tracker	✗	✓	✗	✗	✗
Compositional BMT (mean only)	✗	✗	✓	✗	✗
Compositional BMT	✗	✓	✓	✗	✗
Gaussian Process Regression	✓	✓	✗	✗	✗
Compositional GPR	✓	✓	✓	✗	✗
RL^2	✓	✓	✓	✓	✗
RR-RL^2	✓	✓	✓	✓	✓

comparison of all previously outlined models. We measured the fit to human choices based on two metrics: posterior model frequency and exceedance probability[45]. The posterior model frequency measures how often a model offers the best explanation in the population, while the exceedance probability measures how likely it is that a given model is the most frequent explanation. Further details about this model comparison procedure can be found in the Materials and Methods section.

This model comparison revealed that RR-RL^2 captures how people behave on the first trial of the compositional sub-task the best according to both metrics, with exceedance probability amounting to 0.99, while its posterior model frequency was 0.704 ± 0.002. The compositional BMT is the second-best model with $4.33e-09$ and 0.288 ± 0.002 on exceedance and posterior model frequency respectively. Interestingly, we found that the two models that performed best in our model simulations (compositional GPR and RL^2) did not predict human behaviour well. Taken together, these results support the hypothesis that people do not compose in a fully optimal way, but that their ability to reason compositionally is driven by principles of resource rationality.

We conducted additional model-based analyses to test alternative explanations for sub-optimal zero-shot compositional inference displayed by humans. For instance, we included the corner-arm heuristic model, which picks only the two corner options, and the mean-only variant of the compositional BMT model, which posits people mentally add previously chosen maximally rewarding options. However, the RR-RL^2 model still offered the best explanation for human choices. see the section on "Ruling out alternative hypothesis" in SI for details.

To further support the model comparison results, we simulated behaviour from the two best-fitting models and compared them against human behaviour. With regards to the probability of making the optimal choice on the first trial, we found that simulations from RR-RL^2 ($M = 0.366$, $SE = 0.0154$) matched human behaviour ($M = 0.382$, $SE = 0.021$) whereas the compositional BMT differed significantly ($M = 0.459$, $SE = 0.016$; $t = -2.938$, $p < 0.01$).

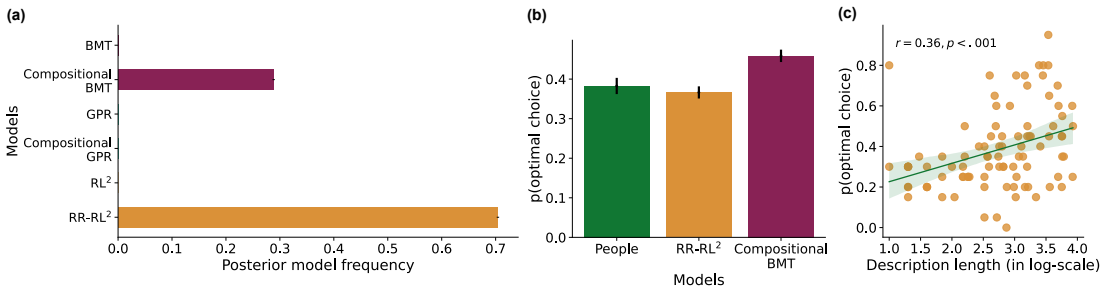

Figure 3. Modelling results of experiment 1. (a) The posterior model frequency of participant choices on the first trial of the last sub-task. (b) Comparison of the probability of making the optimal choice on the first trial of the last sub-task between people and simulations from the best-fitting models: RR-RL^2 and compositional BMT. (c) Correlation between fitted description length of RR-RL^2 (plotted in log-scale) and the probability of making an optimal choice on the first trial of the last sub-task. The fitted regression line is shown in green, with the shaded portion showing the 95% confidence interval.

Next, we examined whether the fitted description lengths of RR-RL^2 could capture task performance. To do this, we correlated the probability of humans making the optimal choice on the first trial against the fitted description lengths for each participant. We found that description length correlated significantly with optimality ($r = 0.359$, $p < 0.001$) as illustrated in Fig. 3(c). Likewise, we also observed a significant negative correlation ($r = -0.36, p < 0.001$) between fitted description lengths and mean regrets on the first trial.

We then analysed whether fitted description lengths can be used to explain the types of choices participants make. For this, we grouped participants based on their fitted description lengths into two groups. In the first group, we included participants whose fitted description lengths were in the range of 1000 to 10000 ($N = 39$), and in the second group, we considered those whose fitted description lengths were in the range of 10 to 100 ($N = 17$). We then compared the probability of making the optimal choice on the first trial of the final sub-task between the two groups. We found that participants in the first group performed near-perfect composition, whereas the choices from participants in the second group were far away from optimality (for further details, see SI). Thus, fitted description lengths can be used to cluster participants into those who can perform near-optimal zero-shot compositional inference and those who cannot.

Additionally, we looked at how regrets on the first two sub-tasks influence behaviour on the first trial in RR-RL^2, just like how we did it in people. We found that the posterior regression coefficients of the model match human behaviour qualitatively with slight deviations. The model picks the non-optimal corner arms and non-optimal phasic arms when it learns one sub-task better than the other (with performance in periodic sub-task having a greater influence on the linear in both cases) and follows a completely different strategy from the ones above when it does not learn both the sub-tasks well. An interesting deviation was that, unlike humans who make more optimal choices when they learn both sub-tasks equally well, the model does so when they learn the linear sub-task better than the periodic sub-task (for detailed results and visualizations, see SI).

Taken together, results from behavioural and model-based analyses suggest that people can perform zero-shot compositional inference but still deviate systematically from optimal behaviour. Their choices are best explained by a meta-reinforcement learning model (RR-RL^2) that learns a solution with limited computational resources. Furthermore, we find the simulated behaviour from RR-RL^2 matches human behaviour well and that description length – the parameter that controls computational resources of RR-RL^2 – correlates with the probability of making an optimal choice on the first trial of the compositional sub-task.[2]

Experiment 2: Change-point rule

Next, we wanted to verify that the results obtained in the previous section transfer to another compositional rule. We, therefore, conducted a second experiment using a change-point rule. The experimental procedure followed the same structure as the additive rule but with participants now being tested on a slot machine whose rewards were composed based on the change-point rule (see Fig. 1 for an example).

Behavioural analysis

Like in the additive rule experiment, we see that people perform much better than chance right from the outset for the curriculum condition ($M = 1.937$, $SE = 0.081$; $t = -15.105, p < 0.001$) while starting at chance-level for the non-curriculum condition ($M = 3.096$, $SE = 0.060$) as shown in Fig. 4(a). We also performed a mixed-effects linear regression analysis as we did in the first experiment which confirmed that participants in the curriculum condition had a significantly lower regret on the first trial of the final sub-task than participants in the non-curriculum condition ($\hat{\beta} = -1.18 \pm 0.115; z = -10.24, p < 0.001$). When looking at the probability of picking the optimal arm in Fig. 4(b), we also find that people make better choices in the curriculum condition ($M = 0.337$, $SE = 0.016$) than in the non-curriculum condition ($M = 0.168$, $SE = 0.008$, $t = -9.347, p < 0.001$). The performance in the curriculum condition is better than in the non-curriculum condition for all trials, which was also the case in the additive rule. Thus, similar to the additive rule experiment, people are able to perform approximate zero-shot compositional inferences.

Even though people were able to compose in a zero-shot manner, they were again not flawless. Their initial regrets in the final sub-task ($M = 1.937$, $SE = 0.081$; $t = 38.25$, $p < 0.001$) deviated significantly from ideal compositional reasoning, thereby corroborating our results from the previous experiment. In addition, people's suboptimality was underlined by the persistent presence of learning effects in the curriculum condition ($\hat{\beta} = -0.216 \pm 0.013; z = -16.112 p < 0.001$).

We also inspected participants' marginal action distribution on the final sub-tasks first trial in Fig. 4(c). We see that the mode of the participants' action distribution lies at the optimal choice. However, human behaviour systematically deviates from optimal behaviour as people pick sub-optimal options frequently. People also tend to pick the corner options – especially the left-most option – quite frequently as they did in the additive rule.

[2]Note that in the main text, we have focused on model-comparison results for the first trial of the compositional sub-task as we are mostly interested in zero-shot compositional inference. However, we also evaluated our models over all trials of the compositional sub-task. The results of these analyses are summarised in the SI.

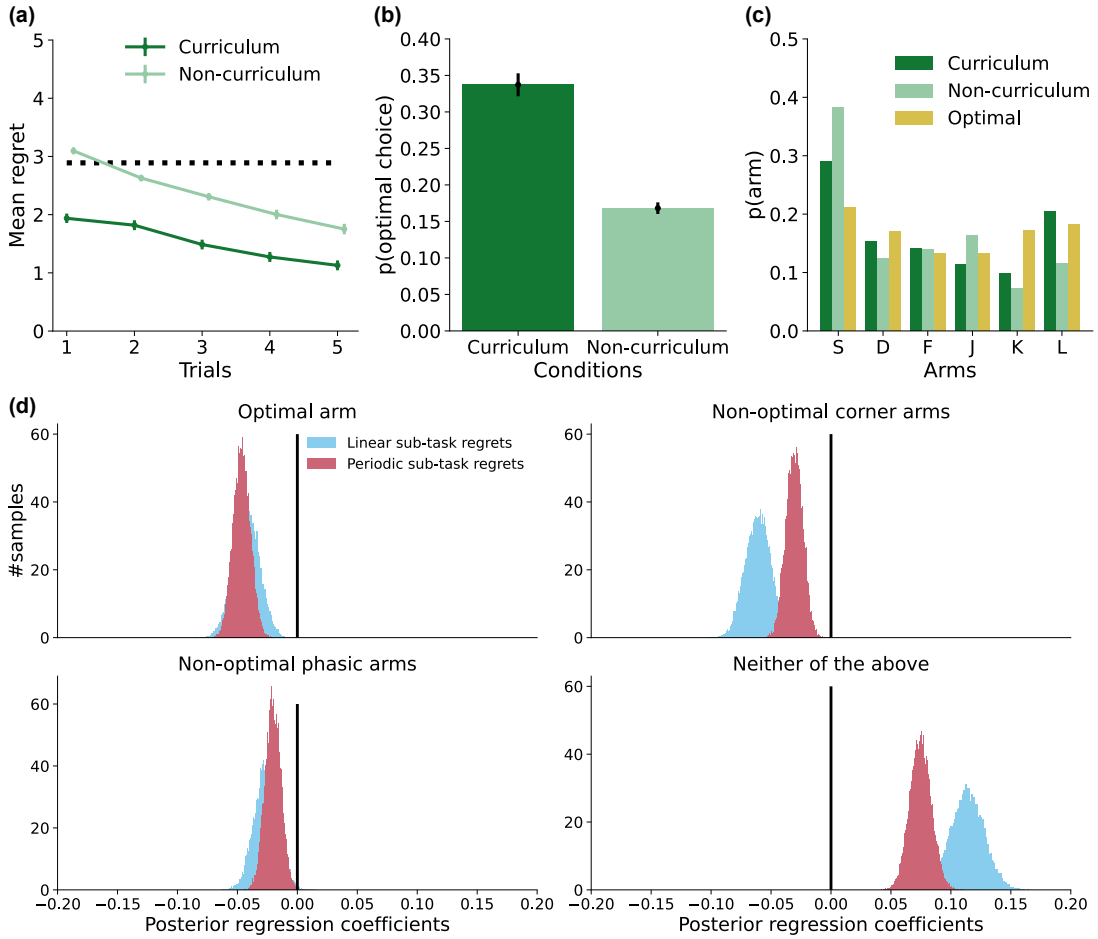

Figure 4. Behavioural results of experiment 2. (a) Mean regrets for participants in the final sub-task for the two conditions: curriculum and non-curriculum. The dotted lines in black indicate mean regret from a random policy. (b) Probability of participants making the optimal choice on the first trial of the final sub-task for the curriculum and non-curriculum condition. Error bars in (a) and (b) represent standard errors computed over participants. (c) Marginal distribution of choices of participants on the first trial of the final sub-task for the two conditions. The bar in gold shows the marginal distribution of choices for the optimal policy. (d) Explaining choices of participants in the last sub-task based on their task performance in the first two sub-tasks. The choices on the first trial were classified into four categories: optimal arm (top-left), non-optimal corner arms (top-right), non-optimal phasic arm (bottom-left), and neither of the above (bottom-right). Then, we fit a Bayesian logistic regression model from the total regrets (summed over all trials) in the linear and periodic sub-tasks onto each of these categories. The sub-plots show the histogram of the posterior regression coefficients of linear and periodic sub-tasks for all four choice categories.

Finally, we repeated the regret analysis we did for the additive rule to better understand which kind of mistakes people make. The results of this analysis are summarized in Fig. 4(d). We see that posterior regression coefficients for regrets of both linear ($M = -0.0423, SE = 7.518e-05$) and periodic ($M = -0.0464, SE = 5.038e-05$) sub-tasks are negative and have overlapping distributions in cases where people picked the optimal option ($t = 45.874, p < 0.001$). This suggests that learning both linear and periodic sub-tasks equally well predicts good performance on the first trial. When people pick non-optimal corner arms, their performance in the linear sub-task ($M = -0.0607, SE = 7.663e-05$) seems to be driving their behaviour more than their performance in the periodic sub-task ($M = -0.0305, SE = 5.0042e-05$). However, on the contrary, picking the non-optimal phasic arm is not driven strongly by periodic sub-task performance with regression coefficients for both linear ($M = -0.0259, SE = 7.0662e-05$) and periodic ($M = -0.0203, SE = 4.672e-05$) sub-tasks overlapping ($t = -66.174, p < 0.001$). Lastly, the posterior regression coefficients are distributed on the positive axis for both linear ($M = 0.114, SE = 9.796e-05$) and periodic ($M = 0.0747, SE = 6.368e-05$) regrets when predicting non-optimal choice belonging to neither to corner or phasic arms.

Model-based analysis

The results for model-based analyses with the change-point rule mirrored that of the additive rule. We find again that RR-RL^2 captures best how people compose according to both metrics with exceedance probability amounting to 0.99, while its posterior model frequency was $0.741 \pm 1.729e-03$. The compositional BMT is the second-best model with an exceedance probability of close to 0 and a posterior model frequency of $0.253 \pm 1.702e-03$ as visualized in Fig. 5(a). These results thus show again that people do not compose in a fully optimal way, but that their ability to reason compositionally is instead impeded by computational constraints.

We simulated behaviour from the two best-fitting models with their parameters fitted to participant behaviour. Looking at the probability of selecting the optimal choice, we found that simulations from RR-RL^2 ($M = 0.2922, SE = 0.0132$) matched human behaviour ($M = 0.3372, SE = 0.0158$) more closely than the compositional BMT ($M = 0.2862, SE = 0.0069$; $t = 2.9560, p < 0.01$) as shown in Fig. 5(b).

We found that fitted description lengths correlated significantly with the probability of participants making an optimal choice for the change-point rule as well, showing a correlation coefficient of $r = 0.487$ ($p < 0.001$) as shown in Fig. 5 (c). The result also holds when we use regrets as a performance measure ($r = -0.448, p < 0.001$).

Following our earlier analysis, we grouped the participants based on their fitted description lengths into two groups with the first group including participants with fitted description lengths between 1000 to 10000 ($N = 58$) and the second group including participants with fitted description lengths between 10 to 100 ($N = 25$). We compared zero-shot compositional inference between the two groups (for detailed analysis, see SI) and found again that participants in the first group performed near-perfect zero-shot compositional inferences, whereas the choices from participants in the second group were far away from optimality.

Finally, we inspected how performance on the first two sub-tasks influenced behaviour on the first trial of the compositional sub-task in RR-RL^2. We found that overall the posterior regression coefficients of the model's behaviour in linear and periodic sub-tasks match human behaviour quite well. They pick the non-optimal corner arms and non-optimal phasic arms when they learn one sub-task better than the other and they follow a completely different strategy from the ones above when they do not learn both the sub-tasks well. However, as in experiment 1 but to a smaller extent, the model makes optimal choices when they learn the linear sub-task better than the periodic sub-task. The results of these analyses are summarised in the SI.

Taken together, these results mirror those we had for the additive rule with similar factors affecting human behaviour in the compositional sub-task. This suggests that people's ability to do compositional inference is robust with regard to the specific way in which functions are composed.

Discussion

Compositionality is at the core of people's ability to generalize from sparse data. It has even been argued to be an essential component of intelligence more generally[5,7,11]. However, how people use this ability to make decisions is less well understood. To address this question, we have proposed a novel experimental paradigm where people first need to learn about latent reward functions in two sub-tasks to be able to pick the most rewarding option on the first trial of the third sub-task. We found that people indeed perform this kind of zero-shot compositional inference, but they deviate systematically from ideal behaviour. Even so, their mistakes were not random but instead highly structured. Extensive model-based analyses revealed that RR-RL^2 – a meta-learned neural network model that accounts for limited computational resources – captures participants' behaviour the best. Mistakes made by this model were also systematic and predicted by similar factors that predicted human suboptimal choices. This result indicates that people seem to follow resource-rational principles when making compositional inferences, thereby expanding on earlier results from other cognitive domains such as decision-making[46], planning[47,48], and problem-solving[49].

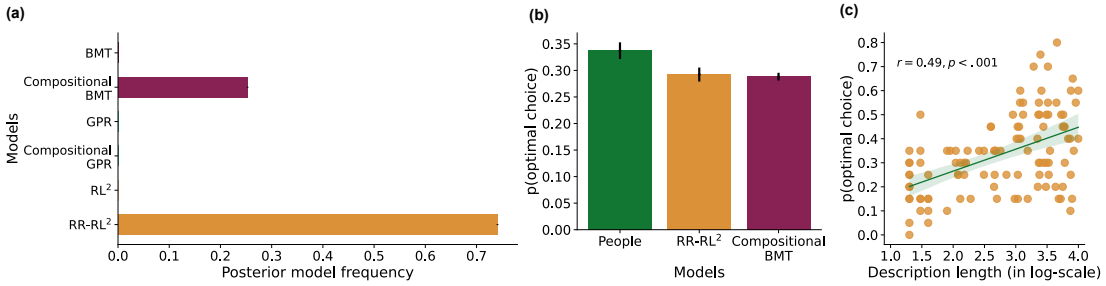

Figure 5. Modelling results of experiment 2. (a) The posterior model frequency of participant choices on the first trial of the last sub-task. (b) Comparison of the probability of making the optimal choice on the first trial of the last sub-task between people and simulations from the best-fitting models: RR-RL^2 and compositional BMT. (c) Correlation between fitted description length of RR-RL^2 (plotted in log-scale) and the probability of making an optimal choice on the first trial of the last sub-task. The fitted regression line is shown in green, with the shaded portion showing the 95% confidence interval.

Relation to empirical works in compositional reinforcement learning

Previous work has investigated how people structure prior knowledge and compose specific components of this learned knowledge to generalize efficiently in reinforcement learning tasks. Xia and Collins[13] have used the options framework from hierarchical reinforcement learning to show that humans can learn hierarchical options and do so such that the temporal ordering of the learned options remains intact. They further showed that people, akin to their model, can compose the learned options to explore novel contexts, thereby speeding up learning. Looking into how learned knowledge guides generalisation, Franklin and Frank[8] showed that human learners decompose learned task structures into distinct components such as rewards and state transitions. They devised a meta-learning agent[50] that trades off re-using these components jointly or compositionally and showed that, similar to this agent, humans too act adaptively on these components depending on the statistics of the task environment. In comparison to our work, these previous studies focused on how people generalize in a sample-efficient way to novel tasks and not on zero-shot compositional inference. For example, Collins and colleagues[10] showed that subjects can transfer previously learned task clusters to a novel context on the first trial, they did not exclusively test for zero-shot compositional inference of previously learned latent concepts as we do. Therefore, our work complements these earlier investigations by addressing a distinct dimension of compositional reasoning. It is worth noting that adapting the proposed models to our experimental paradigm is not straightforward. However, our compositional GPR shares a similar flavour to the hierarchical models proposed by Xia and Collins, indicating potential conceptual connections between different approaches to compositional reasoning.

Model complexity and compositional inference

The relationship between description length and human performance has received considerable attention in previous research[20,23–25]. However, researchers have typically investigated how the description length of the *task* influences performance. For instance, Amalric and colleagues[24] asked participants to predict and repeat sequences displayed on a clock-like display, varying the complexity of the sequences by changing the length of the generating program. They found a correlation between the difficulty of predicting (and repeating) a given sequence and its complexity. In contrast, we investigated how the description length of *strategies* applied by individual participants influenced their performance. We found that different participants apply strategies with different description lengths, implying that they use varying amounts of cognitive resources to perform the task. Thus, task performance is influenced not only by external factors such as task difficulty but also by internal factors such as individual differences in cognitive resources.

Neural networks and compositionality

In contrast to the prevailing notion that neural network models struggle with compositional reasoning[51,52], we found that the model that captured human behaviour best in our experiment was a neural network model, suggesting that these models are not inherently unable to reason compositionally. Instead, it matters how they are set up and how they are trained. This result is supported by other recent studies demonstrating that neural network models can be good models of human compositionality[53–57]. For example, Lake[53] demonstrated that meta-learning can be used to train sequence-to-sequence networks that generalize compositionally in human-like ways on the SCAN data set[51], while Kumar and colleagues[54] found that augmenting meta-reinforcement learning agents with an auxiliary objective to reproduce task descriptions aligns them with human behaviour in a

setting that requires reasoning about compositionally-generated patterns. Finally, Dekker and colleagues[55] designed neural network architecture with an inductive bias for compositional reasoning using a Hebbian gating process, and demonstrated that the resulting model learns composable functions similar to how these functions are learned by people.

Possible neural and cognitive processes

To perform optimal zero-shot compositional reasoning in our task, one must learn the latent reward functions, hold them in working memory until the end of the first two subtasks, and then reason compositionally over these learned latent functions. It is hypothesized that the basal ganglia-striatal circuits, along with the hippocampus, medial temporal cortex, and ventral prefrontal cortex, may be involved in the spatial representation and learning of the latent reward functions underlying the options[58–60]. Maintaining a representation of acquired potential reward functions in memory can involve working memory via prefrontal cortex[61]. In a recent review article, Frankland and colleagues[6] have suggested that – across multiple studies – the default mode network, particularly, the lateral mid-anterior temporal cortex, is engaged in compositional reasoning of learned structures. This suggestion is consistent with previous research that has implicated fronto-cortical and striatal interactions in compositional generalization[10,62]. However, additional research is still necessary to investigate the neural processes that can connect function learning, the maintenance of learned functions in working memory, and their composition. It would be interesting to connect our computational models, in particular RR-RL^2, to some of these results from the neuroscience literature in future work.

Limitations

While one might argue that the proposed compositional bandit task is too simplistic, we found that human behaviour systematically deviated from optimality, implying that the task complexity was appropriate for the investigated research question. Furthermore, our task has two main advantages compared to those previously used to study compositional reasoning. First, it is directly inspired by experiments used to study human learning in structured environments[29], allowing us to connect our findings to previous work on human cognition. The second advantage is its simplistic design. This simplicity allowed us to build computational models that solve the task near-optimally, picking the best option in a zero-shot fashion. Nevertheless, it might be interesting to develop more naturalistic compositional reasoning tasks in future work to test if our model-based predictions still hold. There are additional variants of our paradigm that could be considered. For example, one could test whether increasing the length of the first two sub-tasks causes people to make better compositional inferences. People's performance could furthermore be boosted by relying on a purely observational setting in which options in the first two sub-tasks are presented in a structured manner (for example from left to right).

There are several reasons why participants might be deviating from selecting the optimal arm in the third sub-task. We have investigated factors such as the performance on the first two sub-tasks, learning the values of the options independently without any generalisation, the inability to learn the true underlying basis functions, the inability to explore and compose, and the computational resources to learn in the main text. We furthermore considered several additional explanations such as the corner arm heuristic, mentally adding maximal rewards, etc, in the "Ruling out alternative hypothesis" section placed in the SI. However, there are still factors, for example, forgetting values for certain options, and persisting with function learning strategies from the first two sub-tasks, that could not be taken into account which we leave for future work.

We also note three shortcomings on the modelling side. First, we did not consider resource-rational Bayesian models in our model-based analyses, as building such models is not straightforward. In contrast to this, limited resources are easy to account for in the meta-learning setting[43] which is why we relied on such models instead. Second, our models assume each task to be independent of each other. Learning-to-learn effects could be incorporated into both the Bayesian and the meta-learned models. For the Bayesian models, one could build on the work of Schulz and colleagues[29] who used a simple clustering algorithm to capture the learning-to-learn behaviour across tasks. For meta-learning agents, it would in principle be possible to train over samples of entire experiments (i.e., 20 successive tasks) where each experiment is sampled from a parameterized distribution of experiments. However, training such agents is challenging in practice especially since gradients need to be propagated over longer horizons. Third, the distribution of problems used to train the meta-learning agents in this work does not reflect the statistics of the real world. Recently, Jagadish and colleagues[63] have shown that injecting ecologically valid priors into these agents enables them to explain human category learning better than seven other cognitive models. In future work, it is worth investigating how to design reinforcement learning tasks that capture real-world reward statistics.

Conclusion

We introduced a novel experimental paradigm and two complementary computational approaches for studying zero-shot compositional reinforcement learning in people. We showed that while people can perform zero-shot compositional inference in our task, their choices were better explained by a resource-constrained model than by optimal zero-shot compositional inference. Thus, our results provide a new perspective to the understanding of human compositional inference by considering the influence of cognitive resources. Taken together, our work takes a considerable step towards understanding compositional reinforcement learning in humans, symbolic, and sub-symbolic agents under computational constraints.

Materials and methods

In this section, we provide details of the experimental methods and computational models used to analyse compositional reinforcement learning in humans. In the experimental methods subsection, we describe the task parameters, experimental design, participants, and ethics approval. In the computational methods subsection, we expand on the computational models and explain the methods used to fit these models to human behaviour along with the model comparison procedure.

Experimental methods

Ethics statement

All participants gave informed consent prior to beginning the experiment. Experiments were performed in accordance with the relevant guidelines and regulations approved by the ethic committee of the medical faculty of the University of Tuebingen (Ethik-Kommission an der Medizinischen Fakultät der EberhardKarls-Universität und am Universitätsklinikum Tübingen) with application number 701/2020BO.

In addition, keeping in mind data protection law of Max Planck Institute, we analyzed participants' data anonymously. After agreeing to participate in the study, participants also consented on a data protection sheet approved by the data protection officer of the MPG (Datenschutzbeaufragte der MPG, Max-Planck-Gesellschaf zur Förderung der Wissenschafen).

Participants

We recruited 200 participants (103 female, M_{age} = 28.90) through the Prolific platform for experiment 1.Participants were randomly assigned to the curriculum or non-curriculum condition. All participants had an approval rate of 95% or more, were fluent English speakers from the United States, and were 18 years of age or older. Participants were rewarded a base payment of £2 and a performance-dependent bonus payment up to £2.5. The bonus payment was computed by multiplying the total points earned by the participant with 1e-4. For experiment 2, we recruited 211 participants (96 females, M_{age} = 27.58) through the Prolific platform. The rest of the study parameters remained the same as the additive rule.

Task

Each task consisted of three multi-armed bandit sub-tasks in which rewards follow a function that is dependent on the spatial position of arms:

$$r_t = f(a_t) + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, 0.1) \tag{1}$$

where t denotes the time-step, $a_t \in \{0, \ldots, 5\}$ the arm selected in time-step t and ε_t is an additive noise term. Reward functions f_{linear} and f_{periodic} for the first two sub-tasks are sampled from either the linear or the periodic family as shown below:

$$f_{\text{linear}}(a_t) = \left(\frac{2a_t}{5} - 1\right) w + b + \zeta \qquad w \sim \mathcal{U}(-2.5, 2.5), b \sim \mathcal{U}(2.5, 7.5) \tag{2}$$

$$f_{\text{periodic}}(a_t) = A\,|\sin(0.5\pi\,(a_t - \phi))| + b + \zeta \qquad A \sim \mathcal{U}(0, 7.5), \phi \in \{0, 1\}, b \sim \mathcal{U}\left(0, \frac{A}{1.4}\right) \tag{3}$$

where $\mathcal{U}(a, b)$ is a uniform distribution on the interval [a,b] and $\zeta \sim \mathcal{N}(0, 0.2)$ is an additive noise term. The parameters were chosen after several rounds of piloting to make it easy for participants to perform the task well on average. For example, we found that excluding linear functions with very low slopes and periodic functions with very small amplitudes helped them learn the periodic and linear sub-tasks more easily and hence, improved their performance on the task overall.

Reward functions in the final sub-task were constructed by composing the reward functions encountered in the two earlier sub-tasks. We considered two different composition rules, an additive rule and a change-point rule:

$$f_{\text{additive}}(a_t) = f_{\text{linear}}(a_t) + f_{\text{periodic}}(a_t) \tag{4}$$

$$f_{\text{change-point}}(a_t) = \begin{cases} f_{\text{linear}}(a_t) & \text{if } a_t \in \{0, 1, 2\} \\ f_{\text{periodic}}(a_t) & \text{otherwise} \end{cases} \tag{5}$$

The order of composition in the change-point function was randomized. We set the length of each sub-task to 5 trials, leading to 15 overall trials per task. Note that the number of trials per sub-task was less than the number of available options. This prevents an agent from exhaustively trying out all options and forces it to generalize based on the underlying function.

Experimental design

We conducted an online behavioural study following the structure of the compositional bandit task outlined earlier to test the underlying mechanisms behind how people compose. Participants were told that they were gamblers visiting the fictional town of "Bandit City". They visited multiple casinos (20 in total) in which they played different sets of slot machines. Each casino had two slot machines made by two different companies, called Blue Lagoon and Green Geeks, with their colour (included in the name) indicating the manufacturer. Participants were informed that all slot machines from the same company behaved similarly (i.e. rewards are sampled from the same underlying function, Equation 2 or 3), but were not told which reward function belonged to which company. They had to figure this out via trial and error during the experiment. However, they were shown two canonical examples from each reward function in the task instruction phase to get an idea of how samples from these reward functions could look like. In each casino, participants had five trials per slot machine, with the goal of winning as many coins as possible. In the curriculum condition, they would first interact with a slot machine from each of the two manufacturers. Following their interactions with the two slot machines, participants were tested on a new slot machine, which was a composition of the two previously played machines. Thus, in the curriculum condition, participants interacted with bandits for a total of 300 trials (breakup: 5 trials per sub-task $\times$ 3 sub-tasks $\times$ 20 tasks). Participants assigned to the non-curriculum condition followed the same task structure as the curriculum condition but with minor changes. In this condition, participants were told that the manufacturers only allowed them to play against the compositional slot machine. As a result, participants only interacted with one slot machine with rewards coming from the additive composition which results in a total of 100 trials (breakup: 5 trials per sub-task $\times$ 1 sub-task $\times$ 20 tasks)

Computational models

In this section, we describe the models that can perform the task with each model making a different assumption on how people could be approaching our task. A complete description of the models can be found in the SI.

Bayesian models

Under Gaussian assumptions, a Bayesian Mean-Tracker is often used to track a time-varying reward function $f_t(a)$ for option a on trial t. The mean is assumed to change over trials according to a Gaussian random walk:

$$f_{t+1}(a) \sim \mathcal{N}\left(f_t(a), \sigma_\xi^2\right) \tag{6}$$

where $\sigma_\xi = 0.001$. We also considered a variant of BMT model that can compose learned rewards called compositional BMT. This model follows the same setup as BMT but has its prior mean for the last sub-task initialized to the composition of learned means from the first two sub-tasks. We provide additional details about model learning and inference in the SI.

Gaussian Process Regression models learn a distribution over functions $f(a)$ defined by a mean function $\mu(a)$ and a covariance, or kernel function $k(a,a')$, where a and a' are arms of the bandit. The mean function defines the expected function value, while the covariance function controls the dependence between the function values for different inputs:

$$f(a) \sim \mathcal{GP}\ \ (\mu(a), k(a,a')) \tag{7}$$

$$\mu(a) = \mathbb{E}[f(a)] \tag{8}$$

$$k\left(a,a'\right) = \mathbb{E}\left[\left(f(a)-\mu(a)\right)\left(f\left(a'\right)-\mu\left(a'\right)\right)\right] \tag{9}$$

For the GP-based agent, we considered the GPR model with radial basis function (RBF) as the kernel. k_{RBF} allows the GPR model to generalize its learned value estimates depending on how (spatially) similar the options are to each other.

$$k_{\text{RBF}}\left(a,a'\right) = \exp\left(-\frac{1}{2}\left(a-a'\right)^\top \Theta^{-2}\left(a-a'\right)\right) \tag{10}$$

where Θ is the length scale hyperparameter. Like BMT, we assume that our GPR agent maintains a separate GP for each sub-task and by design, also cannot do any compositional inference.

As GPRs with appropriate priors can approximate the true generative model used for sampling the reward distributions in our task. We constructed a GPR model with compositionality built in, called compositional GPR. Such a model can compose the learned reward estimates from the first two sub-tasks and hence, reason compositionally on the third sub-task. We again assume the agent maintains a separate GP for each sub-task. We set the prior mean of the GPs corresponding to the first two sub-tasks to zero. The covariance function for the first sub-task is defined through a linear kernel k_{linear} defined as

$$k_{\text{linear}}\left(a,a'\right) = va^\top a' \tag{11}$$

where v is the scale hyperparameter. While the covariance function of the second sub-task is defined through a periodic kernel k_{periodic} defined as

$$k_{\text{periodic}}\left(a,a'\right)=\exp\left(-\frac{2\sin^2\left(\pi\mid a-a'\mid/p\right)}{\eta^2}\right) \tag{12}$$

where p is a hyperparameter that determines the period length and η is a lengthscale hyperparameter.

The means and kernels for the final sub-task are obtained by composing the means and kernels from the first two sub-tasks[64]. For the first trial of the additive composition, the kernel is set to the mean of the learned linear and the periodic kernel from the two sub-tasks. The compositional additive kernel k_{additive} is defined as:

$$k_{\text{additive}}\left(a,a'\right)=0.5(k_{\text{linear}}\left(a,a'\right)+k_{\text{periodic}}\left(a,a'\right)) \tag{13}$$

While the prior mean for additive composition is set to the mean of the previously learned mean functions from the linear and periodic sub-tasks. For the first trial of the change-point compositions, the kernel entries are set to that of the linear kernel if both arms belong to the linear function, the periodic kernel if both belong to the periodic function, and zero otherwise. The compositional change-point kernel $k_{\text{change-point}}$ is defined as:

$$k_{\text{change-point}}\left(a,a'\right)=k_{\text{linear}}\left(a,a'\right)\alpha_{\text{linear}}\left(a,a'\right)+k_{\text{periodic}}\left(a,a'\right)\alpha_{\text{periodic}}\left(a,a'\right) \tag{14}$$

where

$$\alpha_{\text{linear}}\left(a,a'\right)=\begin{cases}1 & \text{if } a,a'\in\{0,1,2\}\\ 0 & \text{otherwise}\end{cases} \tag{15}$$

α_{periodic} is defined analogously, giving a value of 1 whenever both arms' $a,a'\in\{3,4,5\}$. Note that we randomized whether the first three arms would belong to the linear or the periodic function. The prior mean in the change-point composition is set to the means learned in linear and periodic sub-tasks for the corresponding arms.

Meta-reinforcement learning

The version of RL^2 we use consists of a recurrent neural network (RNN) network followed by two linear networks that output a policy and a value estimate respectively[39,40]. We denote the joint vector of parameters of this model with $\mathbf{W}$. The network receives task-relevant observations o_t along with the action a_{t-1} and reward from the previous time step r_{t-1} as input and outputs a policy $\pi(a_t|\mathbf{h}_t,\mathbf{W})$ and a value estimate conditioned on the updated hidden state of the RNN. RL^2 is trained on samples from a task distribution $p(\omega)$ to find the policy that maximizes the sum of rewards in an episode of finite horizon H. The full objective function being optimized is shown in Equation 16:

$$\max_{\mathbf{W}}\mathbb{E}_{p(\omega)\prod p(r_t,o_{t+1}|a_t,\omega)\pi(a_t|\mathbf{h}_t,\mathbf{W})}\left[\sum_{t=1}^{H}r_t\right] \tag{16}$$

The particular resource constraint considered in RR-RL^2 is the description length of the meta-learned RL algorithm, which is defined as the number of bits required to store its parameters. Mathematically, this can be accomplished through a simple modification of Equation 16:

$$\begin{aligned}&\max_{\Lambda}\mathbb{E}_{q(\mathbf{W}|\Lambda)p(\omega)\prod p(r_t,o_{t+1}|a_t,\omega)\pi(a_t|h_t,\mathbf{W})}\left[\sum_{t=1}^{H}r_t\right]\\ &\text{s.t. KL}[q(\mathbf{W}|\Lambda)\|p(\mathbf{W})]\leq C\end{aligned} \tag{17}$$

RR-RL^2 differs from RL^2 in two important ways. First, it uses a stochastic parameter encoding over neural network weights $q(\mathbf{W}|\Lambda)$ instead of a point estimate. Second, it places a constraint on the Kullback–Leibler (KL) divergence between $q(\mathbf{W}|\Lambda)$ and a prior $p(\mathbf{W})$, effectively limiting the number of bits that are needed to store the network's parameters and therefore the emerging reinforcement learning algorithm[65].

The network architecture of RL^2 and RR-RL^2 agents consisted of a gated-recurrent unit of size 128[66] followed by two linear layers that map hidden state-to-value function and policy respectively. The model implementations closely followed the implementation of Binz and Schulz[44]. We used a variational dropout prior[67] for RR-RL^2 and assumed that the encoding distribution factorizes into a set of independent normal distributions with learnable means and log-variances. Models were trained using a standard actor-critic loss at the end of each episode[68]. We used the ADAM optimizer[69] with a learning rate of 0.001 and trained for a total of 10^6 episodes with batches of size 32. RR-RL^2 relied on a dual gradient ascent procedure to enforce the constraint on the KL divergence[70]. We obtained gradients w.r.t. the parameters of the encoding distribution λ using the reparametrization trick[71]. We trained RR-RL^2 with description lengths between 10 and 10000 nats.

Modeling fitting and comparison

Bayesian models

The parameters of the Bayesian models were optimised endogenously for each participant, i.e., parameters of the model are chosen to maximise the likelihood of the data observed so far as in Schulz and colleagues[29]. We feed in the choice taken and reward received by participants from the previous trial and predict the expected reward and its uncertainty measure for all six options for the given trial. Note that the predictions are made after each trial conditioned on all data points up to that trial in the given task. The kernel parameters of these models are learned via gradient descent using the ADAM optimiser[69] for 100 iterations. The initial prior noise of these models was set to 0.001.

Following prior work[72,73], we use a variant of upper confidence bound sampling with an additional stickiness component as an action selection policy for both Bayesian models:

$$z(a_t|\beta,\tau,\lambda) = \boldsymbol{\beta}\boldsymbol{\mu}(a_t) + \boldsymbol{\sigma}(a_t) + \boldsymbol{\delta}(a_t,a_{t-1}) \tag{18}$$

where $\delta(a_t,a_{t-1})$ takes the value of 1 if $a_t = a_{t-1}$ and 0 otherwise. This formulation includes uncertainty estimates for the learned values as an additional term to guide exploration. It has been shown to capture human behaviour well in function learning tasks[28,73] and also comes with performance guarantees[74].

The policy $p_{\text{Bayesian}}(a_t|\beta,\tau,\lambda)$ is then derived from these values using the softmax function:

$$p_{\text{Bayesian}}(i) = \frac{e^{z(a_t)}}{\sum_{j=1}^{K} e^{z(a_t)}} \quad for\ a_t = 0,1,\ldots,5 \tag{19}$$

The free parameters β,τ, and λ were fitted to human choices using a Bayesian model fitting procedure for each participant separately with the priors for parameters set to $\mathcal{N}(0,5)$. Model fitting was performed using the probabilistic programming toolbox PYMC3[75]. We used the marginal likelihood on the first trial of the compositional sub-task for model comparisons.

Meta-reinforcement learning

For modelling human choices, we assumed a mixture policy of the policy provided by the meta-reinforcement learning agent, a random policy, and a stickiness term:

$$p_{\text{RL}^2}(a_t \mid \varepsilon,\lambda) = (1-\varepsilon-\lambda)\pi(a_t \mid \mathbf{h}_t) + \varepsilon|\mathscr{A}|^{-1} + \boldsymbol{\delta}(a_t,a_{t-1}) \tag{20}$$

$$p_{\text{RR-RL}^2}(a_t \mid \varepsilon,\lambda,C) = (1-\varepsilon-\lambda)\pi(a_t \mid \mathbf{h}_t,C) + \varepsilon|\mathscr{A}|^{-1} + \boldsymbol{\delta}(a_t,a_{t-1}) \tag{21}$$

where C, ε and λ are free parameters, $|\mathscr{A}|$ denotes the number of available actions, and $\delta(a_t,a_{t-1})$ takes the value of 1 if $a_t = a_{t-1}$ and 0 otherwise. The marginal distribution $\pi(a_t \mid \mathbf{h}_t)$ was approximated with 10 samples from the encoding distribution.

We performed a grid search over the free parameters ε, λ and C and obtained a log-likelihood estimate for all pairs of parameters. ε and λ could take values between 0 and 1 with increments of 0.02, subject to the constraint that their sum is less than or equal to 1. The description length C could take values from 10 to 10,000 in steps of 10. We assumed a uniform prior probability over these discretized parameter values, which allows us to compute the marginal log-likelihood for the first trial of the compositional sub-task as follows:

$$\log \sum_{\varepsilon}\sum_{\lambda}\sum_{C} \exp\left(\sum_{n=1}^{N} p_{\text{RR-RL}^2}(a_{11,n,i} \mid \varepsilon,\lambda,C)\right) - \log(N_C \cdot N_\varepsilon \cdot N_\lambda) \tag{22}$$

where N_C, N_ε, and N_λ correspond to the number of considered values for each parameter.

Model comparison

To obtain a quantitative measure of the goodness-of-fit to human choices, we conducted a Bayesian model comparison of all previously outlined models. We provide the full list of fitted parameters for each model in the SI. We measured the fit to human choices based on two metrics: posterior model frequency and exceedance probability[45]. The posterior model frequency measures how often a model offers the best explanation in the population, while the exceedance probability measures how likely it is that a given model is the most frequent explanation. We compute the metrics for model comparison using a Python implementation of the Variational Bayesian Analysis (VBA) toolbox [URL]. The toolbox requires us to provide log evidence – the marginal log-likelihood from the model fitting procedure in our case – for each model and participant, which we compute as previously described. For further details about this model comparison procedure see Rigoux and colleagues[45].

Supporting information

S1 Text. Supplementary information file including fine-grained behavioural analysis, implementation details for the different models and the model comparison procedure, additional model comparison results, other alternative hypothesis that were ruled out, confidentiality statement, compute and data/code availability. This file includes Figure 1 (Regrets in linear and periodic sub-tasks), Figure 2 (Mean regrets for linear and periodic sub-tasks across tasks), Figure 3 (Number of unique options picked by participants over tasks), Figure 4 (Extent of zero-shot compositional inference.), Figure 5 (Analysis of regrets for same and different condition in experiment 1), Figure 6 (Analysis of regrets for same and different condition in experiment 2), Figure 7 (Model simulations, Figure 8 (Exceedance probability), Figure 9 (Model comparison for all trials of the last sub-task.), Figure 10 (Model comparison for first trial last sub-task), Figure 11 (RR-RL^2 model with low description lengths display corner-arm bias), Figure 12 (Comparing performance of participants better fitted by Bayesian Mean Tracker with those by RR-RL^2), Figure 13 (Marginal log-likelihoods over trials), Figure 14 (Composers versus non-composers), and Figure 15 (Analysis of RR-RL^2 regrets). Figure legends see inside S1.

Acknowledgements

This work was supported by the Max Planck Society, the Volkswagen Foundation, the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A, and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Germany Excellence Strategy-EXC2064 / 1–390727645.

Author contributions statement

A.J., M.B., and E.S. conceived the experiments, A.J. conducted the experiments, A.J. and T.S. analysed the results under M.B.'s, J.W.'s, and E.S.'s supervision. A.J. and M.B. wrote the main draft of the manuscript and J.W., T.S., and E.S. provided comments and reviewed the manuscript.

References

1. Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. How to grow a mind: Statistics, structure, and abstraction. *Science* **331**, 1279–1285 (2011).
2. Carey, S. & Bartlett, E. Acquiring a single new word. *Reports on Child Lang. Dev.* (1978).
3. Schmidt, L. A. *Meaning and compositionality as statistical induction of categories and constraints*. Ph.D. thesis, Massachusetts Institute of Technology (2009).
4. Griffiths, T. L. Revealing ontological commitments by magic. *Cognition* **136**, 43–48 (2015).
5. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *arXiv preprint arXiv:1604.00289* (2016).
6. Frankland, S. M. & Greene, J. D. Concepts and compositionality: in search of the brain's language of thought. *Annu. review psychology* **71**, 273–303 (2020).
7. Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).
8. Franklin, N. T. & Frank, M. J. Generalizing to generalize: Humans flexibly switch between compositional and conjunctive structures during reinforcement learning. *PLoS computational biology* **16**, e1007720 (2020).
9. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. review* **94**, 115 (1987).
10. Collins, A. G. E. & Frank, M. J. Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning. *Cognition* **152**, 160–169 (2016).
11. James, W. *The Principles Of Psychology Volume II By William James (1890)* (Henry Holt and company, New York, 1890).
12. Kemp, C. Exploring the conceptual universe. *Psychol. Rev.* **119**, 685 (2012).
13. Xia, L. & Collins, A. G. Temporal and state abstractions for efficient learning, transfer, and composition in humans. *Psychol. review* **128**, 643 (2021).
14. von Humboldt, W. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistigen Entwickelung des Menschengeschlechts* (Königliche Akademie der Wissenschaften, 1836).
15. Chomsky, N. *Aspects of the Theory of Syntax*, vol. 11 (MIT Press, 2014).

16. Griffiths, T. L. & Tenenbaum, J. B. Theory-based causal induction. *Psychol. review* **116**, 661 (2009).

17. Griffiths, T. L. Formalizing prior knowledge in causal induction. *The Oxf. Handb. Causal Reason.* 115 (2017).

18. Schulz, E. *Towards a unifying theory of generalization*. Ph.D. thesis, UCL (University College London) (2017).

19. Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M. & Gershman, S. J. Probing the compositionality of intuitive functions. Tech. Rep., Center for Brains, Minds and Machines (CBMM) (2016).

20. Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M. & Gershman, S. J. Compositional inductive biases in function learning. *Cogn. Psychol.* **99**, 44–79, DOI: 10.1016/j.cogpsych.2017.11.002 (2017).

21. Sanborn, A. & Griffiths, T. L. Markov chain monte carlo with people. In *Advances in Neural Information Processing Systems*, 1265–1272 (2008).

22. Duvenaud, D., Lloyd, J. R., Grosse, R., Tenenbaum, J. B. & Ghahramani, Z. Structure discovery in nonparametric regression through compositional kernel search. In *Proceedings of the 30th International Conference on Machine Learning*, 1166–1174 (2013).

23. Kumar, S., Dasgupta, I., Cohen, J., Daw, N. & Griffiths, T. Meta-learning of structured task distributions in humans and machines. In *International Conference on Learning Representations* (2020).

24. Amalric, M. *et al.* The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS computational biology* **13**, e1005273 (2017).

25. Piantadosi, S. T., Tenenbaum, J. B. & Goodman, N. D. The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychol. review* **123**, 392 (2016).

26. Sablé-Meyer, M., Ellis, K., Tenenbaum, J. & Dehaene, S. A language of thought for the mental representation of geometric shapes. *Cogn. Psychol.* **139**, 101527 (2022).

27. Planton, S. *et al.* A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. *PLoS computational biology* **17**, e1008598 (2021).

28. Schulz, E., Konstantinidis, E. & Speekenbrink, M. Putting bandits into context: How function learning supports decision making. *J. experimental psychology: learning, memory, cognition* **44**, 927 (2018).

29. Schulz, E., Franklin, N. T. & Gershman, S. J. Finding structure in multi-armed bandits. *Cogn. Psychol.* **119**, 101261 (2020).

30. Saanum, T., Schulz, E. & Speekenbrink, M. Compositional generalization in multi-armed bandits. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43 (2021).

31. Jagadish, A. K., Saanum, T., Wang, J. X., Binz, M. & Schulz, E. Probing compositional inference in natural and artificial agents. In *5th Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM 2022)*, 275–279 (2022).

32. Cole, M. W., Laurent, P. & Stocco, A. Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control. *Cogn. Affect. & Behav. Neurosci.* **13**, 1–22 (2013).

33. Cole, M. W. *The biological basis of rapid instructed task learning*. Ph.D. thesis, University of Pittsburgh (2009).

34. Lieder, F. & Griffiths, T. L. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* **43** (2020).

35. Gershman, S. J., Horvitz, E. J. & Tenenbaum, J. B. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science* **349**, 273–278 (2015).

36. Speekenbrink, M. & Konstantinidis, E. Uncertainty and exploration in a restless bandit problem. *Top. cognitive science* **7**, 351–367 (2015).

37. Rasmussen, C. E. Gaussian processes in machine learning. In *Summer school on machine learning*, 63–71 (Springer, 2003).

38. Schulz, E., Speekenbrink, M. & Krause, A. A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **85**, 1–16 (2018).

39. Duan, Y. *et al.* RI^2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779* (2016).

40. Wang, J. X. *et al.* Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763* (2016).

41. Schubert, J. A., Jagadish, A. K., Binz, M. & Schulz, E. In-context learning agents are asymmetric belief updaters. *arXiv preprint arXiv:2402.03969* (2024).

42. Ortega, P. A. *et al.* Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030* (2019).

43. Binz, M. *et al.* Meta-learned models of cognition. *arXiv preprint arXiv:2304.06729* (2023).

44. Binz, M. & Schulz, E. Modeling human exploration through resource-rational reinforcement learning. *Adv. Neural Inf. Process. Syst.* **35**, 31755–31768 (2022).

45. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies—revisited. *Neuroimage* **84**, 971–985 (2014).

46. Bhui, R., Lai, L. & Gershman, S. J. Resource-rational decision making. *Curr. Opin. Behav. Sci.* **41**, 15–21 (2021).

47. Correa, C. G., Ho, M. K., Callaway, F., Daw, N. D. & Griffiths, T. L. Humans decompose tasks by trading off utility and computational cost. *arXiv preprint arXiv:2211.03890* (2022).

48. Callaway, F. *et al.* Rational use of cognitive resources in human planning. *Nat. Hum. Behav.* **6**, 1112–1125 (2022).

49. Binz, M. & Schulz, E. Reconstructing the einstellung effect. *Comput. Brain & Behav.* 1–17 (2022).

50. Franklin, N. T. & Frank, M. J. Compositional clustering in task structure learning. *PLoS computational biology* **14**, e1006116 (2018).

51. Lake, B. & Baroni, M. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In *International conference on machine learning*, 2873–2882 (PMLR, 2018).

52. Hupkes, D., Dankers, V., Mul, M. & Bruni, E. Compositionality decomposed: How do neural networks generalise? *J. Artif. Intell. Res.* **67**, 757–795 (2020).

53. Lake, B. M. Compositional generalization through meta sequence-to-sequence learning. *Adv. neural information processing systems* **32** (2019).

54. Kumar, S. *et al.* Using natural language and program abstractions to instill human inductive biases in machines. *Adv. Neural Inf. Process. Syst.* **35**, 167–180 (2022).

55. Dekker, R. B., Otto, F. & Summerfield, C. Determinants of human compositional generalization. *PsyArXiv preprint* (2022).

56. Peng, X. B., Chang, M., Zhang, G., Abbeel, P. & Levine, S. Mcp: Learning composable hierarchical control with multiplicative compositional policies. *Adv. Neural Inf. Process. Syst.* **32** (2019).

57. Andreas, J., Klein, D. & Levine, S. Learning with latent language. *arXiv preprint arXiv:1711.00482* (2017).

58. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464–1468 (2016).

59. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. neuroscience* **20**, 1643–1653 (2017).

60. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267–279 (2014).

61. Collins, A. G. & Frank, M. J. How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* **35**, 1024–1035 (2012).

62. Collins, A. G. & Frank, M. J. Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol. review* **120**, 190 (2013).

63. Jagadish, A. K., Coda-Forno, J., Thalmann, M., Schulz, E. & Binz, M. Ecologically rational meta-learned inference explains human category learning. *arXiv preprint arXiv:2402.01821* (2024).

64. Duvenaud, D., Lloyd, J., Grosse, R., Tenenbaum, J. & Zoubin, G. Structure discovery in nonparametric regression through compositional kernel search. In *International Conference on Machine Learning* (2013).

65. Hinton, G. E. & Van Camp, D. Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the sixth annual conference on Computational learning theory*, 5–13 (1993).

66. Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

67. Kingma, D. P., Salimans, T. & Welling, M. Variational dropout and the local reparameterization trick. *Adv. neural information processing systems* **28** (2015).

68. Mnih, V. *et al.* Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, 1928–1937 (2016).

69. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

70. Haarnoja, T. *et al.* Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905* (2018).

71. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

72. Borji, A. & Itti, L. Bayesian optimization explains human active search. *Adv. neural information processing systems* **26** (2013).

73. Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D. & Meder, B. Generalization guides human exploration in vast decision spaces. *Nat. human behaviour* **2**, 915–924 (2018).

74. Srinivas, N., Krause, A., Kakade, S. M. & Seeger, M. W. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE transactions on information theory* **58**, 3250–3265 (2012).

75. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in python using pymc3. *PeerJ Comput. Sci.* **2**, e55 (2016).

Supplementary Information: A resource-rational account of zero-shot compositional inference in a reinforcement learning setting

Akshay K. Jagadish[1,2,3,*,+], Marcel Binz[1], Tankred Saanum[1,3], Jane X. Wang[4], and Eric Schulz[1]

[1]Institute for Human-Centered AI, Helmholtz Computational Health Center, Munich, Germany
[2]University of Tübingen, Germany
[3]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[4]Google Deepmind, London, United Kingdom
[*]akshaykjagadish@gmail.com
[+]Current Address: Institute for Human-Centered AI, Helmholtz Computational Health Center, Munich, Bavaria, Germany

1 Fine-grained behavioural analysis

Performance of participants in the first two sub-tasks: In the main paper, we reported how participants' performance in the first two sub-tasks can explain their behaviour on the final compositional sub-task. In Supplementary Fig. 1, we show the participant's regrets in the linear and periodic sub-task of the curriculum condition in experiment 1 and experiment 2. We found that participants continue to learn over trials in these two sub-tasks as well. Thereby, they lend support to the hypothesis tested in the main text that performance in the final sub-task could be affected by how well participants' have learned the latent functions in the first two sub-tasks.

Zero-shot compositional inference when the most rewarding option is same across sub-tasks than when it is different: We divided the final sub-task in the curriculum condition into two groups. One in which the optimal choices are the same across all three sub-tasks, which we call the "same" condition, and the other in which the optimal choice in the third sub-task differs from the first two sub-tasks, which we call the "different" condition. In Figure 4, we compared the probability of picking the optimal option and regret on the first trial of the third sub-task for both conditions. We found that the mean p(optimal) is higher for the "same" condition ($M = 0.5086, SD = 0.0311$) than the "different" condition ($M = 0.3406, SD = 0.0233$), with the difference being significant ($t = 6.4946, p < 0.0001$). However, their performance in the "different" condition is still significantly better than the non-curriculum condition ($M = 0.1921, SD = 0.00732, t = 3.2499, p < 0.01$). The same results hold for experiment 2 as well. The mean p(optimal) is once again significantly higher for the "same" condition ($M = 0.3553, SD = 0.0171$) than the "different" condition ($M = 0.2826, SD = 0.0245, t = 3.1182, p < 0.01$), but the latter is still significantly better than the non-curriculum condition ($M = 0.1681, SD = 0.0080, t = 2.5552, p < 0.05$). These results indicate that people can perform zero-shot compositional reasoning in our task.

Next, we repeated the logistic regression analyses for the two groups. We found that for the "same" condition in experiment 1, linear and periodic regrets predicted the optimal performance on the first trial equally well. In the "different" group, linear and periodic regrets had a much stronger influence on predicting the optimal arm. The other sub-plots follow the same trend as the original analysis, with linear regrets and periodic regrets predicting non-optimal corner and phasic choices on the first trial more strongly than their counterparts respectively. For details, refer to the main text. In Figure 5 and 6, we show the results from the analysis of regrets for the same and the different conditions for experiment 1 and experiment 2 respectively.

Learning to learn across tasks: To investigate whether participants get the task only after an "Aha" moment, we first tested if there is any significant difference in mean regret performance, across participants, between tasks and found that there were no such differences. We also found no significant differences in the zero-shot compositional inference, when averaged over all participants. Therefore, we fitted a regression discontinuity design (RDD) model independently to each participant to test if they showed a sharp change in task performance in the experiment. This is done by introducing a variable, called "treatment effect" into a linear regression model that regresses task indices onto task performance with an intercept. The treatment effect is assigned a value of one if greater than the threshold otherwise it is assigned a value of zero. The threshold value for the treatment effect is determined by maximum likelihood estimation. We compared the RDD and linear regression model using Bayesian Information Criterion (BIC). As reported in Table 1, we found that the linear regression model fitted human data

better for all sub-tasks in experiment 1 and experiment 2. This indicates that participants did not have an explicit 'Aha moment' whilst learning in our task.

Table 1. Comparing Bayesian information criterion (BIC) – lower the better – between linear regression (LR) model and regression discontinuity design (RDD) for explaining 'Aha moment'.

	Experiment 1		Experiment 2	
Subtask	**LR**	**RDD**	**LR**	**RDD**
Linear	3914.9622	4239.0604	4489.8277	4850.1920
Periodic	5494.4120	5765.8175	6011.9035	6330.1861
Composition	6844.2816	7181.2735	7247.4654	7622.3551

2 Models

2.1 Bayesian Mean-Tracker

If $\mathscr{D}_{t-1}$ encapsulates the history of actions and rewards taken until trial $t-1$, the posterior distribution of the reward function is a Gaussian distribution.

$$P(f_t(a) \mid \mathscr{D}_{t-1}) = \mathscr{N}(f_t(a); m_t(a), v_t(a)) \tag{1}$$

where mean m_t is defined as:

$$m_t(a) = m_{t-1}(a) + \delta_{t-1}(a) G_{t-1}(a) [r_{t-1} - m_{t-1}(j)] \tag{2}$$

and variance as:

$$v_t(a) = [1 - \delta_{t-1}(a) G_{t-1}(a)] \left[v_{t-1}(j) + \sigma_\zeta^2 \right] \tag{3}$$

where $\delta_{t-1}(a)$=1 if arm a was chosen on trial $t-1$ or 0 otherwise. Where gain G_t defined as :

$$G_t(a) = \frac{v_t(a) + \sigma_\zeta^2}{v_t(a) + \sigma_\zeta^2 + \sigma_\varepsilon^2} \tag{4}$$

can be seen as analogous to the learning rate.

2.2 Gaussian Process Regression

The posterior distribution over functions after observing a given data set of observations is also a GP with the mean vector μ_{post} and covariance matrix $\mathbf{K}_{\text{post}}$ defined as follows:

$$\mu_{\text{post}} = \mathbf{K}_{\mathbf{a},\mathbf{a}^*}^\top \left(\mathbf{K}_{\mathbf{a},\mathbf{a}} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{r} \tag{5}$$

$$\mathbf{K}_{\text{post}} = \mathbf{K}_{\mathbf{a}^*,\mathbf{a}^*} - \mathbf{K}_{\mathbf{a},\mathbf{a}^*}^\top \left(\mathbf{K}_{\mathbf{a},\mathbf{a}} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{K}_{\mathbf{a},\mathbf{a}^*} \tag{6}$$

$$\sigma_{\text{post}}(a) = \mathbf{K}_{\text{post}}(a,a) \tag{7}$$

where $\mathbf{r}$ is a vector of rewards for chosen arms, $\mathbf{K}_{\mathbf{a},\mathbf{a}^*}$ is the covariance matrix between previously chosen arms $\mathbf{a}$ and all other arms $\mathbf{a}^*$, $\mathbf{K}_{\mathbf{a},\mathbf{a}}$ is the covariance matrix between all previously chosen arms, and $\mathbf{K}_{\mathbf{a}^*,\mathbf{a}^*}$ is the covariance matrix between all arms the agent can pick.

2.3 Model simulation

We simulated all six models mentioned in the main text on our compositional bandit task and measured their performance in terms of regrets. In Supplementary Fig. 7, we show the regrets on the final sub-task. We considered two policies for the Bayesian models during the model simulation phase: an upper confidence bound policy[1,2] and an ε-greedy policy. We performed a grid search procedure to determine the optimal hyperparameter values for the policies. We report results from the agent with ε-greedy policy as it achieved higher performance on the task. For the meta-reinforcement learning models, we

performed a single forward pass through the converged meta-learned agents on the compositional bandit task for a total of 1000 simulation runs and sampled from the resulting policies. Network parameters were kept constant during this stage.

We found that performance on the first trial was substantially above chance level for four of the models that can compose (the compositional BMT and GPR, as well as the two meta-learning agents), indicating that they can re-use the earlier learned functions to compose new functions in a zero-shot manner. The highest performance was obtained by RL^2, followed by the compositional GPR and BMT models. RR-RL^2's performance was meditated by its description length, with high description length leading to models that matched the performance of RL^2. In contrast to this, the non-compositional agents (BMT and GPR) started the final sub-task by performing at the chance level and continued to get better over trials. However, their performance does not reach the same level as compositional agents even after the stipulated number of trials.

2.4 Analysis of errors

As mentioned in the main text, we fitted a separate Bayesian logistic regression model in PyMC3 from the total regrets (summed over all trials) in the first two sub-tasks onto four choice categories (optimal arm, non-optimal corner arms, non-optimal phasic arm, and neither of the above) coded as a binary variable. The regression coefficients for the total regrets in the first two sub-tasks were estimated for each participant separately using Normal priors, $\mathcal{N}(0,1)$, for both coefficients. For details about the exact Markov chain Monte Carlo sampling method we used for parameter estimation, please refer to the project repository linked below.

3 Model comparison

The full list of fitted parameters for each model is shown in Table 2. The terms β and τ correspond to coefficients for the mean and uncertainty predictions made by the Bayesian models on a trial-by-trial basis. C and ε correspond to the description length and ε of the ε-greedy policy used by the meta-reinforcement learning agents. The λ corresponds to the coefficient towards the stickiness to the option picked one time-step before and is common to both types of models.

Table 2. Fitted parameters in each model.

Model	Parameters
Bayesian Mean-Tracker	β, τ, λ
Compositional Bayesian Mean-Tracker	β, τ, λ
Gaussian Process Regression	β, τ, λ
Compositional Gaussian Process Regression	β, τ, λ
RL^2	ε, λ
RR-RL^2	ε, λ, C

Model comparison results for the first trial: In Supplementary Fig. 8, we show that RR-RL^2 best captures how people compose on the first trial of the last sub-task with exceedance probability amounting to approximately 0.99 in both experiment 1 and experiment 2.

Model comparison over all trials: We also additionally evaluated our models' capacity to predict human choices over all trials of the compositional sub-task. To do this, we repeated our Bayesian model comparison procedure but now for model predictions for all trials of the final sub-task. Such a comparison tells us how different models fare in terms of capturing the *learning-to-compose* behaviour within the last task. We show two performance measures (a-b) posterior model frequency (PMF) and (b-c) exceedance probability (EP) in Supplementary Fig. 9 with results for experiment 1 on the left-column and experiment 2 on the right column. We found that compositional Bayesian mean-tracker (EP=0.3680, PMF=0.3165) comes a close second to RR-RL^2 (EP=0.4833, PMF=0.3299) in terms of both performance measures in experiment 1. Whereas in experiment 2, the Bayesian mean-tracker (EP=1.0, PMF=0.6241) clearly outperforms RR-RL^2 (EP=0., PMF=0.1819).

We were particularly surprised by how well the Bayesian mean-tracker and its compositional variant captured human behaviour in our tasks as these models learn values for each option independently without any generalisation across options Thus, we conducted a more fine-grained analysis as a follow-up where we compare the marginal log-likelihood (MLL) of the top two competing models on a trial-by-trial basis, as shown in Supplementary Fig. 13. We found that although RR-RL^2 explains the human choices the best on the first trial in both experiments, compositional BMT and BMT can capture quite a lot of variance when all trials in the final task are taken into account in experiment 1 and experiment 2. The difference between RR-RL^2 and the compositional BMT in experiment 1 is not significant based on an independent two-sample t-test for any of the trials in experiment 1. Whereas between RR-RL^2 and the BMT in experiment 2, the differences are significant

for all except trial 2 with BMT performing better than RR-RL^2. This could be because RR-RL^2 models do not display any learning to compose behaviour over trials but rather just compose to different extents depending on their description length as shown in Supplementary Fig. 7. Whereas, the Bayesian mean-tracker and its compositional variant fitted to human choices continue to learn over trials. They learn value estimates for all the different options people have picked so far, i.e. all trials until the given trial, which can then be successfully used to predict future human choices. This could also partly be attributed to the model fitting procedure we used from[3], which involves optimising the model parameters endogenously for each participant (see Methods section for details). Following this procedure results in the model placing a high probability mass on all choices participants have made in the experiment until the given trial. Conversely, other models, including the RL^2, learn the true underlying reward function over options and do not place high probability mass precisely on previously chosen options. The by-product of this artefact manifests in the trial-by-trial comparison of the marginal log-likelihoods shown in Supplementary Figure 13, where compositional BMT demonstrates a good fit to human choices by the end of the sub-task. However, considering how sticky people were (i.e. their tendency to pick the last picked choice) in our task, the Bayesian mean-tracker and its compositional variant capturing decent variance in human data should not come as a surprise.

Participants showing zero-shot transfer are better fit by RR-RL^2: We tested whether subjects showing evidence of zero-shot transfer are better fit by the RR-RL^2 than by the BMT model. To do this, we first split participants into two groups based on the marginal log-likelihoods, which was derived by fitting the model to participants' choices across all trials of the final sub-task. In Figure 12, we show the regrets on the compositional sub-task between the groups across all trials (a & c) and on the first trial (b & c). We found the mean regret of RR-RL^2 participants ($M = 1.7903, SE = 0.0977$) was significantly lower than that of the BMT participants ($M = 2.6077, SE = 0.0967$) in experiment 1 ($t = 5.9563, p < 0.000001$). Similarly, in experiment 2 ($t = 6.9343, p < 0.000001$), the mean regret RR-RL^2 participants ($M = 1.4064, SE = 0.0943$) was lower than that of the BMT participants ($M = 2.1482, SE = 0.0505$). Taken together, these results indicate that participants better fitted by RR-RL^2 display better zero-shot transfer than those by BMT.

Composers versus non-composers: To supplement the analysis in the main text, we performed an analysis where we first grouped the participants based on their fitted description lengths into two groups with the first group including participants with fitted description lengths between 1000 to 10000 and the second group including participants with fitted description lengths between 10 to 100. Then, we compared the zero-shot compositional inference between the two groups and showed the results in Supplementary Fig. 14 for experiment 1 and experiment 2. We found that participants in the first group, labelled as composers, performed near-perfect zero-shot compositional inferences in both experiments 1 and 2. Whereas the choices from participants in the second group labeled non-composers, were far away from optimality in both experiments 1 and 2 and behaved similarly to participants from non-curriculum conditions.

Analysis of regrets of RR-RL^2 model choices: We inspected how the performance of RR-RL^2 in the first two sub-tasks influences the behaviour on the first trial of the compositional sub-task and show the results from this analysis in Supplementary Fig. 15. We see that the posterior regression coefficients match human behaviour qualitatively, especially in experiment 2. They also do so in experiment 1 but with slight discrepancies. We found that the agents make optimal choices when they learn the linear sub-task better than the periodic sub-task in both experiment 1 and experiment 2, unlike humans where they do so when they learn both sub-tasks equally well. Additionally, we also found that agents also seem to be picking the non-optimal arms when they learn periodic sub-tasks better than linear sub-tasks in experiment 1. However, the rest of the results are in order with agents in both experiments following a completely different strategy from the ones above when they do not learn both the sub-tasks well.

4 Ruling out alternative hypothesis

Mentally adding maximal rewards from memory: We added a new model to the model comparisons to test the hypothesis that people mentally add numbers from memory and then choose the maximum from the sums. This model, which we call "Compositional Bayesian Mean-tracker (mean only)", composes learned means from the first two sub-tasks without considering their uncertainties. Therefore, it disregards the exploration component during function learning. The model comparison which included this new model still showed that the resource rational meta-reinforcement learning model (RR-RL^2) still explains choices on the first trial better than other models in both experiments($M = 0.5049, SE = 2.6593e-03$). However, the posterior model frequency of the "Compositional Bayesian Mean-tracker (mean only)" model ($M = 0.0809, SE = 7.9126e-04$) is almost the same as that of the "Compositional Bayesian Mean-tracker" ($M = 0.0807, SE = 7.8957e-04$) in experiment 1. In experiment 2, similar results were found with the posterior model frequency of the "Compositional Bayesian Mean-tracker (mean only)" model ($M = 0.06437, SE = 5.4260e-04$) is almost the same as that of the "Compositional Bayesian Mean-tracker" ($M = .00645, SE = 5.4339e-04$). For details, see Figure 10.

Corner-arm heuristic: In experiment 1, we found that the zero-shot compositional inference on compositions where corner options are the best choices across all three sub-tasks was better ($M = 0.5086, SE = 0.0311$) than when they were not ($M = 0.3406, SE = 0.0233; t = 6.4946, p < 0.0001$). However, their performance on the non-corner options is still better than the non-curriculum condition ($M = 0.1921, SE = 0.00732; t = 3.2499, p < 0.01$) indicating that they can perform zero-shot compositional reasoning, albeit not optimally, see Figure 4 (a).

We additionally included a new model, called the corner-arm heuristic, into the model comparison whose policy is a mixture of picking the corner options, a random policy, and a stickiness term:

$$p_{\text{Corner}}(a_t \mid \varepsilon, \lambda) = (1 - \varepsilon - \lambda)\pi(a_t) + \varepsilon|\mathscr{A}|^{-1} + \lambda\,\delta(a_t, a_{t-1}) \tag{8}$$

where C, ε and λ are free parameters, $|\mathscr{A}|$ denotes the number of available actions, and $\delta(a_t, a_{t-1})$ takes the value of 1 if $a_t = a_{t-1}$ and 0 otherwise. The policy $\pi(a_t)$ is to pick the two corner options with equal probability.

The new model comparison showed that RR-RL^2 is still the best model ($M = 0.5049, SE = 2.6593e-03$) even though the corner-arm heuristic does explain people's choices to an extent in terms of posterior model frequency ($M = 0.2111, SE = 1.7718e-03$). In terms of exceedance probability, RR-RL^2 explains human data better than other models. Furthermore, it can be seen in Figure 11 that the RR-RL^2 model with low description lengths (in maroon) shows the same tendency to pick corner options as participants. These results indicate that RR-RL^2 accounts for the various ways in which participants experience difficulty in mentally overlaying the rewards.

Reward-based learning versus verification: We compared the mean number of selected options per sub-task across all participants against the total number of trials within a sub-task to test if participants are engaged in verification and testing. If participants choose all options in every sub-task, it indicates that they might be engaging in verifying the function. However, we found that the mean number of options selected was significantly less than five and continues to decrease across tasks, see Figure 3. Furthermore, we found that the mean regrets decrease significantly across trials (within each sub-task) and tasks in both experiments which suggests that they are actively trying to learn the underlying latent function to earn more points, see Figure 2. Lastly, it is important to note that participants were exclusively instructed to maximize rewards and had to pass an attention check which verifies if they have fully understood the task instruction.

5 Confidentiality, Participation/Withdrawal, and Data Protection Statements

Instructions as given to the participants at the beginning of the experiment: "Your participation in this study will remain confidential. Your Prolific ID will NOT be shared with anyone outside the research team. Your participation in this study is completely voluntary and you may refuse to participate or you may choose to withdraw at any time. However, you will only be paid for participation upon completion and if you enter the correct completion code provided at the end of the study. Your data will be made anonymous and only used in the manner described in our data protection sheet, available here." For more details, check out the human experiment directory within the project's GitHub repository mentioned below.

6 Compute

We used an OpenHPC-based in-house cluster to run all analyses. Specifically, we used a CPU-only cluster with 28 nodes, 512 GB (DDR4 3200 MHz) memory, and running Ubuntu on AMD EPYC 7452 with 64 cores. Each model – one single seed – took about 4-5 hours to train.

7 Data and code availability

We make the experimental data, and the code-base used to run the experiment, analyze participants' behaviour, and perform model-based analysis available on GitHub via this repository:
https://github.com/akjagadish/resource-rational-compositional-RL

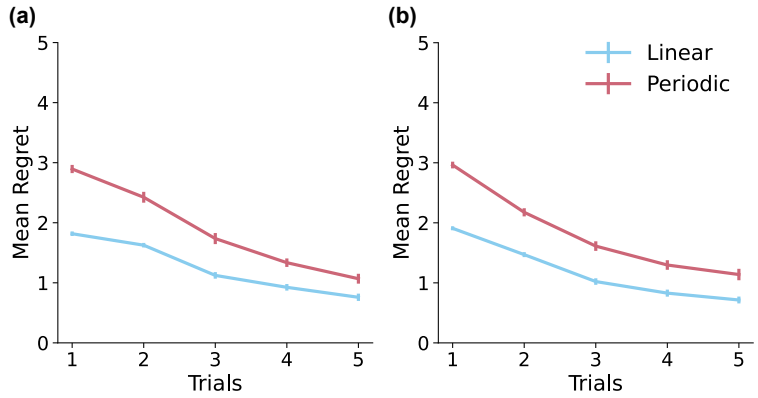

Figure 1. Regrets in linear and periodic sub-tasks. (a-b) Mean regrets for participants in the linear (in orange) and periodic (in blue) sub-task in (a) experiment 1 and (b) experiment 2. Error bars represent standard errors computed over participants.

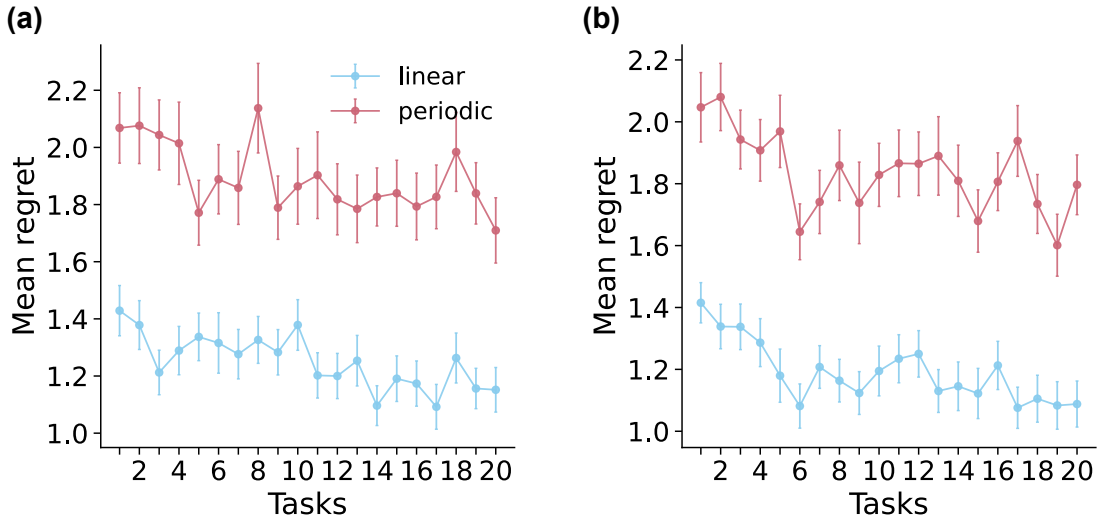

Figure 2. Mean regrets for linear and periodic sub-tasks across tasks. (a-b) Mean regrets over the twenty tasks for participants in the linear (in orange) and periodic (in blue) sub-task in (a) experiment 1 and (b) experiment 2. When task indices are linearly regressed onto tasks regrets, the fitted slope for linear ($\hat{\beta} = -0.1129 \pm 0.030, p < 0.0001$) and periodic ($\hat{\beta} = -0.1159 \pm 0.046, p < 0.05$) sub-tasks are negative and significant in experiment 1. Similarly, in experiment 2, the fitted slope for linear ($\hat{\beta} = -0.1159 \pm 0.027, p < 0.0001$) and periodic ($\hat{\beta} = -0.1147 \pm 0.039, p < 0.005$) sub-tasks are negative and significant. Error bars represent standard errors computed over participants.

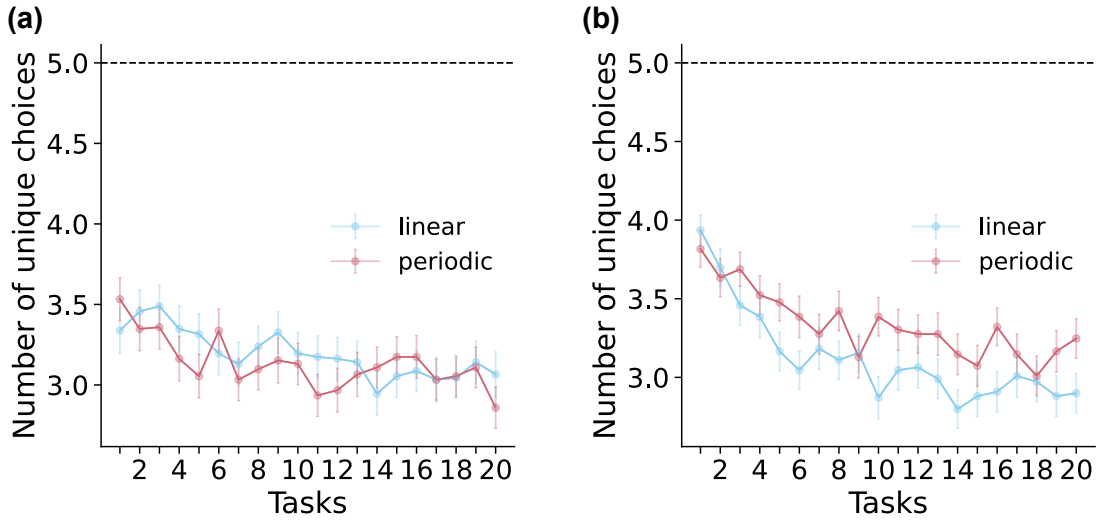

Figure 3. Number of unique options picked by participants over tasks. For samples from linear (in blue) and periodic (in red) latent reward functions for (a) experiment 1 and (b) experiment 2. When task indices are linearly regressed onto tasks regrets, the fitted slope for linear ($\hat{\beta} = -0.1956 \pm 0.049, p < 0.0001$) and periodic ($\hat{\beta} = -0.1690 \pm 0.048, p < 0.001$) sub-tasks are negative and significant in experiment 1. Similarly, in experiment 2, the fitted slope for linear ($\hat{\beta} = -0.3819 \pm 0.047, p < 0.0001$) and periodic ($\hat{\beta} = -0.2831 \pm 0.046, p < 0.0001$) sub-tasks are negative and significant. Error bars represent standard errors computed over participants.

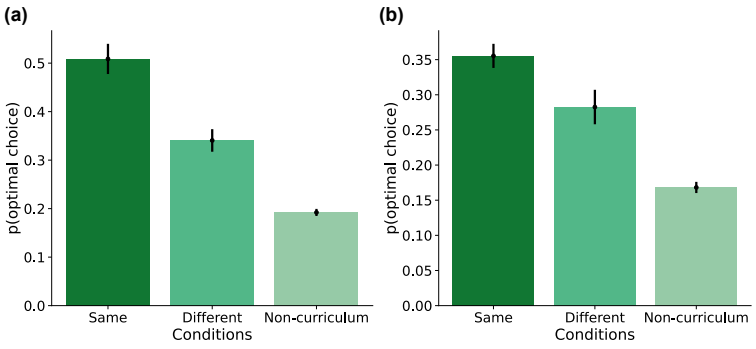

Figure 4. Extent of zero-shot compositional inference. Probability of participants making the optimal choice on the first trial of the final sub-task in (a) experiment 1 and (b) experiment 2. We divided the final sub-task in the curriculum condition into two groups. One in which the optimal choices are the same across all three sub-tasks, which we call the "Same" condition, and the other in which the optimal choice in the third sub-task differs from the first two sub-tasks, which we call the "Different" condition. Error bars represent standard errors computed over participants

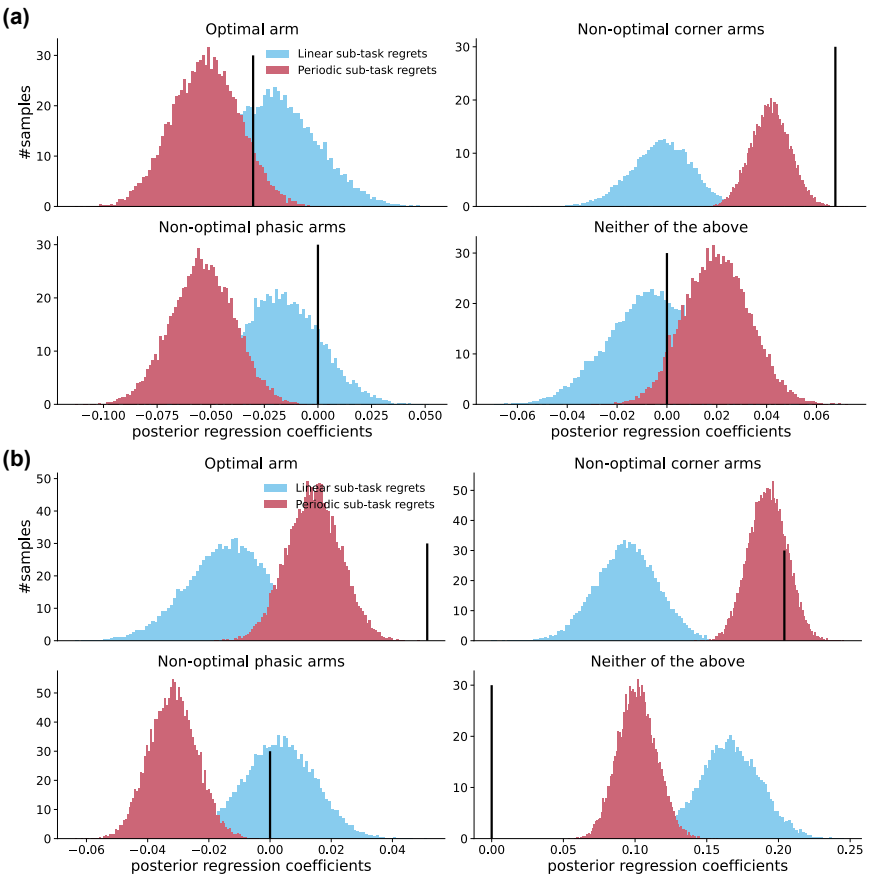

Figure 5. Analysis of regrets for experiment 1: (a-b) Explaining the performance of humans in the last sub-task based on their task performance in the first two sub-tasks for (a) same and (b) different conditions. The same procedure illustrated in the main text for explaining human task performance was followed.
(a) When the optimal arm was picked, periodic sub-task ($M = -0.0191, SE = 0.0001$ $t = 163.1679, p < 0.001$) had a lower mean than linear sub-task ($M = 0.0068, SE = 0.0001$); When the agent picked the non-optimal corner arms, their performance in linear sub-task ($M = -0.1867, SE = 0.0002$; $t = -424.7110, p < 0.001$) was better than the periodic sub-task ($M = -0.0696, SE = 0.0001$). But for the non-optimal phasic arm, the periodic sub-task ($M = -0.05428, SE = 0.0001$; $t = 211.8507, p < 0.001$) was better than the linear sub-task ($M = -0.01807, SE = 0.0001$) as expected. For neither, both linear ($M = -0.0052, SE = 0.0001$) and periodic sub-tasks ($M = 0.0200, SE = 0.0001$) performance were worse. (b) For optimal choices, linear sub-task ($M = -0.0639, SE = 0.0001$; $t = -245.9030, p < 0.001$) had a lower mean than periodic sub-task ($M = -0.0364, SE = 0.0001$); When agent picked the non-optimal corner arms, their performance in linear sub-task ($M = -0.0638, SE = 0.0001$; $t = -551.774, p < 0.001$) was slightly better than the periodic sub-task ($M = -0.0066, SE = 0.0001$). Whereas, for the non-optimal phasic arm, the linear sub-task ($M = 0.0027, SE = 0.0001$; $t = 342.900, p < 0.001$) was better as expected than the periodic sub-task ($M = -0.0318, SE = 0.00005$). Lastly, when neither of the options mentioned above were picked both linear ($M = 0.1665, SE = 0.0001$) and periodic sub-tasks ($M = 0.1014, SE = 0.0001$) performance were worse.

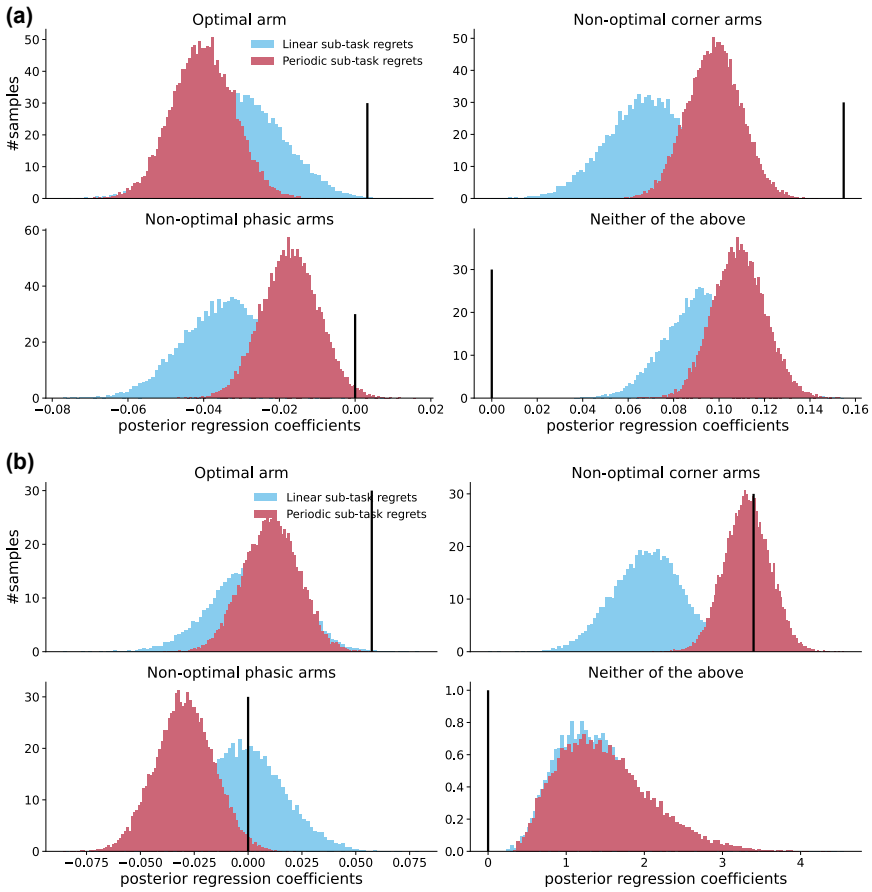

Figure 6. Analysis of regrets for experiment 2: (a-b) Explaining the performance of humans in the last sub-task based on their task performance in the first two sub-tasks for (a) same and (b) different conditions. The same procedure illustrated in the main text for explaining human task performance was followed.
(a) When the optimal arm was picked, linear sub-task ($M = -0.0346, SE = 0.0001$; $t = 83.666, p < 0.001$) had a lower mean than periodic sub-task ($M = -0.0433, SE = 0.0001$); When the agent picked the non-optimal corner arms, their performance in linear sub-task ($M = -0.0600, SE = 0.0001$; $t = -189.1949, p < 0.001$) was slightly better than the periodic sub-task ($M = -0.03976, SE = 0.0001$). But for the non-optimal phasic arm, the linear sub-task ($M = -0.0337, SE = 0.0001$; $t = -171.923, p < 0.001$) was better than the periodic sub-task ($M = -0.0171, SE = 0.00005$) as expected. For neither, both linear ($M = 0.09380, SE = 0.0001$) and periodic sub-tasks ($M = 0.1089, SE = 0.0001$) performance were worse as expected. (b) For optimal choices, linear sub-task ($M = -0.0678, SE = 0.0002$; $t = -47.3271, p < 0.001$) had a lower mean than periodic sub-task ($M = -0.0583, SE = 0.0001$); When agent picked the non-optimal corner arms, their performance in linear sub-task ($M = -0.0636, SE = 0.0001$; $t = -339.8967, p < 0.001$) was slightly better than the periodic sub-task ($M = -0.0038, SE = 0.0001$). Whereas, for the non-optimal phasic arm, the linear sub-task ($M = -0.0036, SE = 0.0001$; $t = 154.274, p < 0.001$) was better as expected than the periodic sub-task ($M = -0.0296, SE = 0.0001$). Lastly, when neither of the options mentioned above were picked, both linear ($M = 1.4199, SE = 0.0039$) and periodic sub-tasks ($M = 1.49851, SE = 0.0042$) performance were worse.

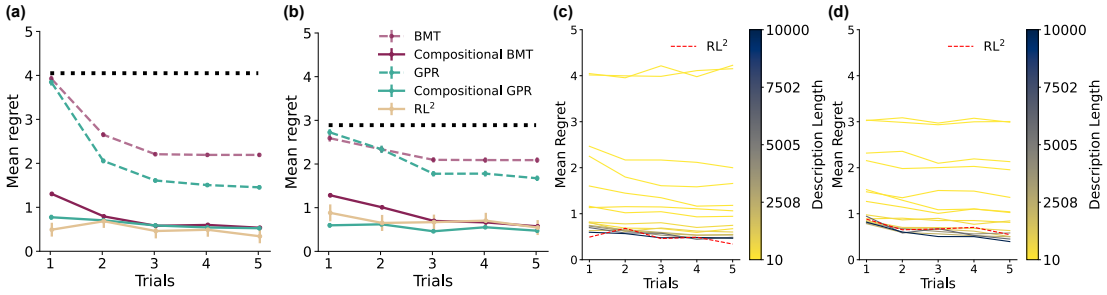

Figure 7. Model simulations. (a-b) Mean regrets for the models averaged over simulation runs in the final sub-task in (a) experiment 1 and (b) experiment 2. The dotted lines in black indicate mean regret from a random policy. (c-d) Mean regrets of RR-RL^2 averaged over simulation runs in the final sub-task for different description lengths in (c) experiment 1 and (d) experiment 2. 35 different description lengths were sampled uniformly within the interval 10 to 10000 (in log-space). The mean regret of RL^2 (in red) was included to show that it lower bounds the performance of RR-RL^2 on the same task. Error bars for models show standard errors over 1000 simulation runs

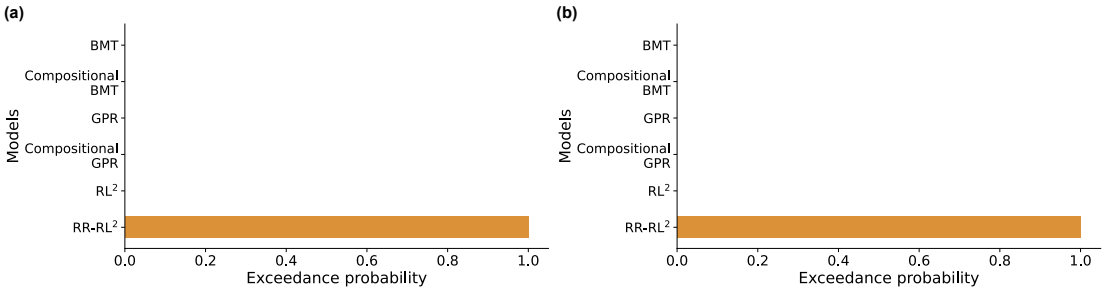

Figure 8. Exceedance probability. (a-b) Exceedance probabilities for all competing models fitted to participant choices on the first trial of the last sub-task in (a) experiment 1 and (b) experiment 2.

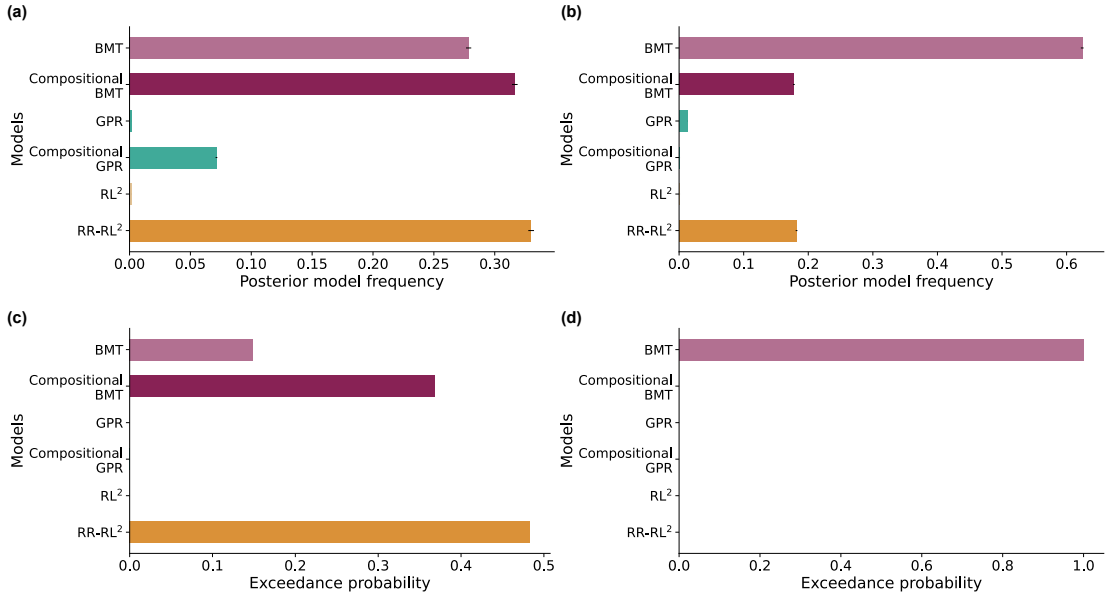

Figure 9. Model comparison for all trials of the last sub-task. (a-b) The posterior model frequency of all models fitted to participant choices over all trials of the last sub-task in (a) experiment 1 and (b) experiment 2. (c-d) Exceedance probabilities for all competing models fitted to participant choices on all trials of the last sub-task in (c) experiment 1 and (d) experiment 2.

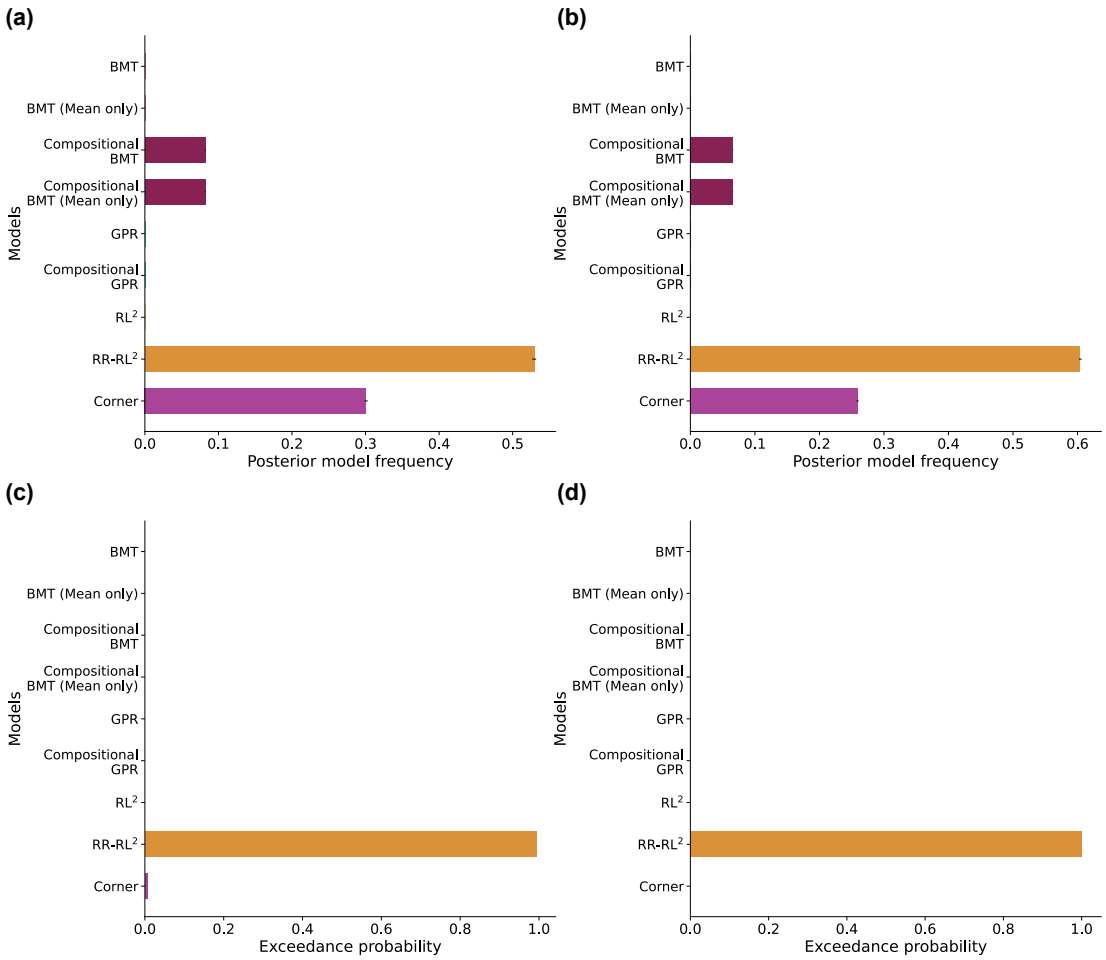

Figure 10. Model comparison for first trial last sub-task. (a-b) The posterior model frequency of all models fitted to participant choices on the first trial of the last sub-task in (a) experiment 1 and (b) experiment 2. (c-d) Exceedance probabilities for all competing models fitted to participant choices on the first trial of the last sub-task in (c) experiment 1 and (d) experiment 2.

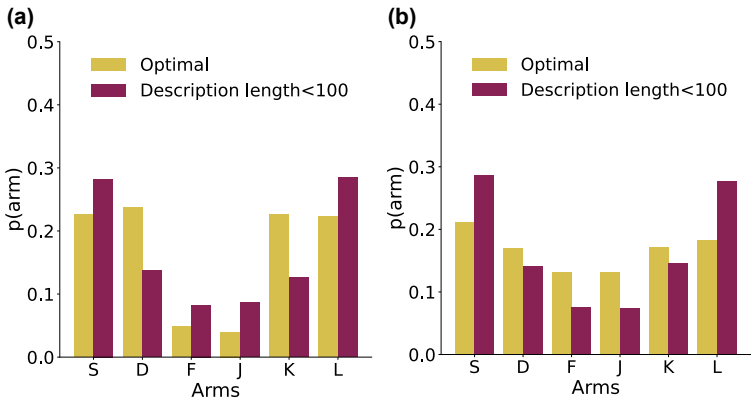

Figure 11. RR-RL^2 model with low description lengths display corner-arm bias. (a-b) Marginal distribution of simulated choices from RR-RL^2 model on the first trial of the final for (a) experiment 1 and (b) experiment 2. The simulated choices from RR-RL^2 model with description lengths in the range of 10 to 100 were pooled together and compared against optimal choices.

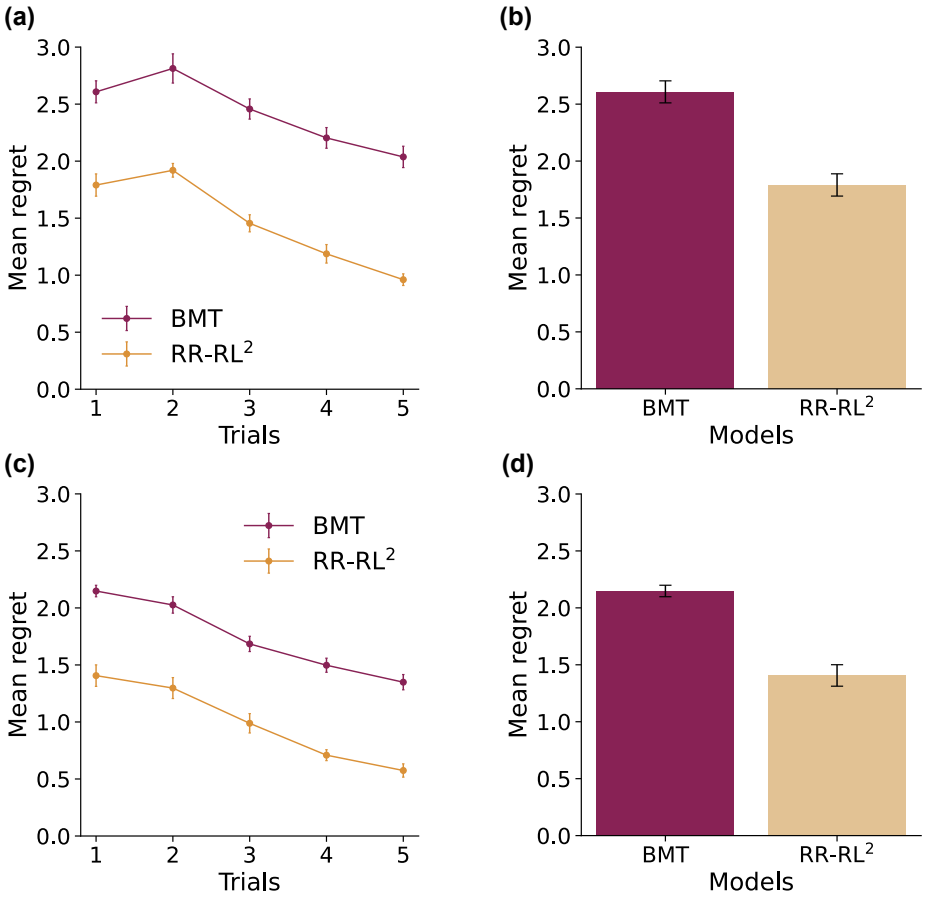

Figure 12. Comparing performance of participants better fitted by Bayesian Mean Tracker with those by RR-RL^2. (a & c) Mean regrets over trials for participants in the final sub-task in (a) experiment 1 and (c) experiment 2. (b & d) Regrets on the first trial for participants in the final sub-task in (b) experiment 1 and (d) experiment 2. BMT refers to the Bayesian Mean Tracker model with upper confidence bound-based sampling policy and RR-RL^2 refers to the resource-rational meta-reinforcement learning model. Participants are grouped into two groups: the BMT group refers to participants who are better explained by the BMT model and RR-RL^2 group consists of those who are better fit by the RR-RL^2 model.

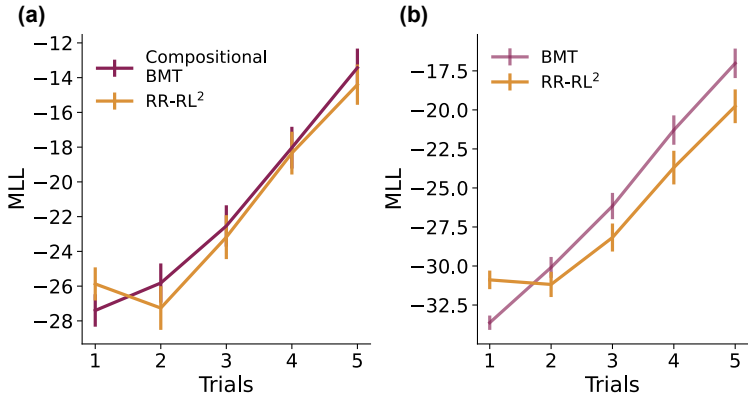

Figure 13. Marginal log-likelihoods over trials. (a) Marginal log-likelihood (MLL) of the top two competing models: Compositional Bayesian Mean-tracker (BMT) and RR-RL^2 fitted to human choices in experiment 1 on a trial-by-trial basis. (b) Marginal log-likelihood of the top two competing models in experiment 2: Bayesian Mean-tracker (BMT) and RR-RL^2.

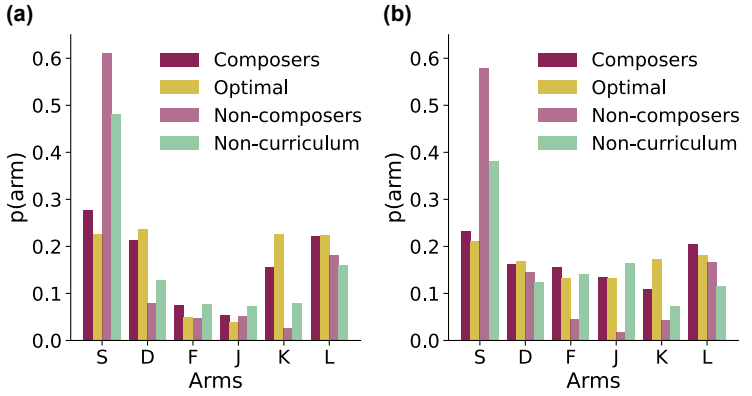

Figure 14. Composers versus non-composers. (a-b) Probability of making the optimal choice on the first trial of the final sub-task between composers, non-composers, optimal and non-curriculum conditions for (a) experiment 1 and (b) experiment 2. Participants were put into two groups based on their fitted description lengths from the RR-RL^2 model. In the high description length group (called, the composers), participants whose fitted description lengths were in the range of 1000 to 10000 were included and in the low description group (referred to as the non-composers), we considered those whose fitted description lengths were in the range of 10 to 100.

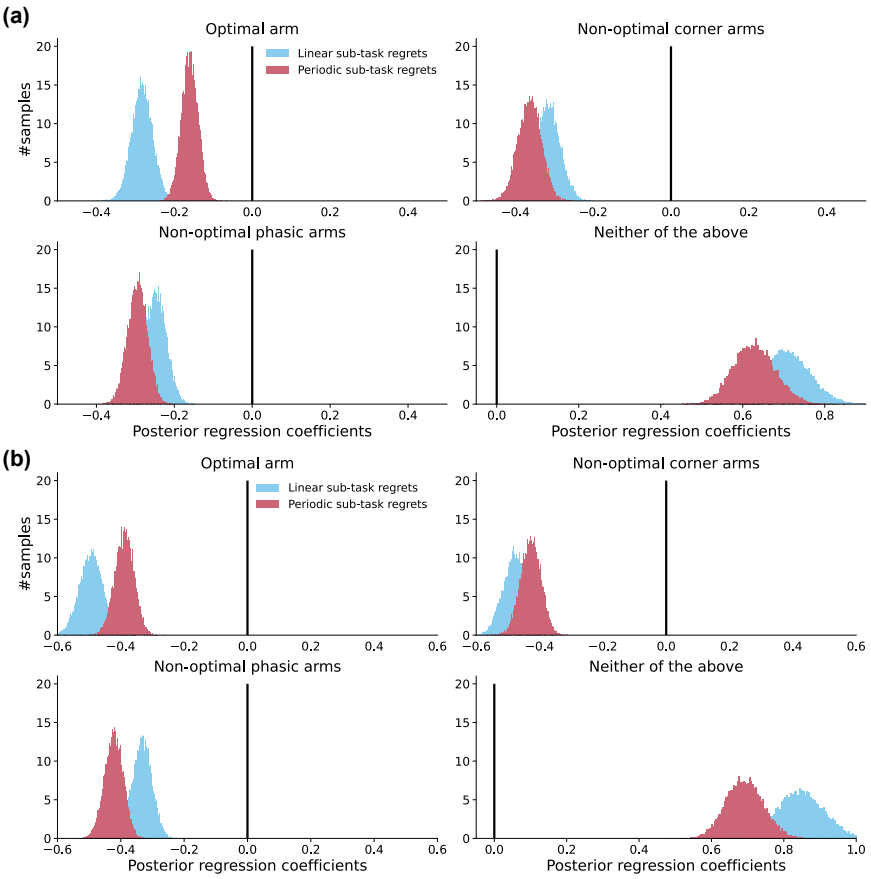

Figure 15. Analysis of RR-RL^2 regrets: (a-b) Explaining the performance of RR-RL^2 in the last sub-task based on its task performance in the first two sub-tasks for (a) experiment 1 and (b) experiment 2. The same procedure illustrated in the main text for explaining humans task performance was followed.
(a) When the optimal arm was picked, linear sub-task ($M = -0.2830, SE = 0.0002$; $t = -505.0662, p < 0.001$) had a lower mean than periodic sub-task ($M = -0.1603, SE = 0.0001$); When the agent picked the non-optimal corner arms, their performance in periodic sub-task ($M = -0.3629, SE = 0.0002$; $t = 146.6408, p < 0.001$) was slightly better than the linear sub-task ($M = -0.3174, SE = 0.0002$). But for the non-optimal phasic arm, the periodic sub-task ($M = -0.2946, SE = 0.0002$; $t = 173.6793, p < 0.001$) was better than the linear sub-task ($M = -0.24828, SE = 0.0002$) as expected. For neither, both linear ($M = 0.7048, SE = 0.0004$) and periodic sub-tasks ($M = 0.6280, SE = 0.0004$) performance were worse as expected. (b) For optimal choices, linear sub-task ($M = -0.4940, SE = 0.0003$; $t = -314.1053, p < 0.001$) had a lower mean than periodic sub-task ($M = -0.3886, SE = 0.0002$); When agent picked the non-optimal corner arms, their performance in linear sub-task ($M = -0.4773, SE = 0.0003$; $t = -141.5513, p < 0.001$) was slightly better than the periodic sub-task ($M = -0.4280, SE = 0.0002$). Whereas, for the non-optimal phasic arm, the periodic sub-task ($M = -0.4221, SE = 0.0002$; $t = 293.7915, p < 0.001$) was better as expected than the linear sub-task ($M = -0.3328, SE = 0.0002$). Lastly, when neither of the above-mentioned options were picked, both linear ($M = 0.8469, SE = 0.0004$) and periodic sub-tasks ($M = 0.6918, SE = 0.0004$) performance were worse.

References

1. Borji, A. & Itti, L. Bayesian optimization explains human active search. *Adv. neural information processing systems* **26** (2013).

2. Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D. & Meder, B. Generalization guides human exploration in vast decision spaces. *Nat. human behaviour* **2**, 915–924 (2018).

3. Schulz, E., Franklin, N. T. & Gershman, S. J. Finding structure in multi-armed bandits. *Cogn. Psychol.* **119**, 101261 (2020).

2.4 Human-like category learning by injecting ecological priors from large language models into neural networks

Jagadish, A. K., Coda-Forno, J., Thalmann, M., Schulz, E.*, & Binz, M.* (2024). Human-like category learning by injecting ecological priors from large language models into neural networks. In Forty-first International Conference on Machine Learning. doi:10.5555/3692070.3692919. arXiv:2402.01821.

Contributions in-context

Ecological rationality, introduced by Todd and Gigerenzer [8], proposes that the human mind is adapted to its surrounding environment through the use of simple, context-specific strategies, called heuristics (see Section 1.7 in Background for details). However, constructing models of ecological rationality has been challenging for two reasons: (1) designing ecologically valid tasks is difficult [130, 131], and (2) even with access to such tasks, it remains difficult to build computational models that can effectively solve them by extracting useful shared structure (see Section 2.1 in Publications).

In this work, we addressed both challenges. To tackle the first, we demonstrated that LLMs can be used to generate ecologically valid classification tasks at scale. We achieved this by querying LLMs to synthesize realistic task features, which was followed by generating values for those features. We verified the ecological validity of the resulting tasks by comparing their data distributional properties to those of the classification tasks from a machine learning benchmarking suite, OpenML-CC18 [132]. Across multiple dimensions, the LLM-generated tasks closely matched the statistical properties of real-world problems.

To address the second challenge, we derived learning algorithms adapted to the statistics of the LLM-generated tasks using meta-learning. Specifically, by substituting real-world classification tasks, which are insufficient in quantity to support meta-learning from scratch, with large volumes of tasks generated by LLMs, we were able to derive ecologically rational learning algorithms. We refer to the resulting class of models as ecologically rational meta-learned inference (ERMI).

We then demonstrated that ERMI not only explains human choices better than seven other domain-specific models, but also captures key qualitative aspects of human category learning: (1) it found the same category structures difficult as humans did [14]; (2) it became more exemplar-based as learning progressed, mirroring humans [133]; (3) it displayed human-like patterns of generalization when tested on novel stimuli. In addition, ERMI achieved state-of-the-art performance on the OpenML-CC18 classification benchmark [132].

Finally, we discussed the limitations of this framework, from biases in the data to missing cognitive constraints and the challenge of scaling the approach to derive a multi-modal meta-learned model (see Section 3.1 and Section 3.2 in Outlook for an extended discussion). We concluded

with a provocative question: How much variance in human behavior can a rational model adapted to ecologically valid problems explain?

Human-like Category Learning by Injecting Ecological Priors from Large Language Models into Neural Networks

Akshay K. Jagadish [1 2] **Julian Coda-Forno** [1 2] **Mirko Thalmann** [1 2] **Eric Schulz** [*1 2] **Marcel Binz** [*1 2]

Abstract

Ecological rationality refers to the notion that humans are rational agents adapted to their environment. However, testing this theory remains challenging due to two reasons: the difficulty in defining what tasks are ecologically valid and building rational models for these tasks. In this work, we demonstrate that large language models can generate cognitive tasks, specifically category learning tasks, that match the statistics of real-world tasks, thereby addressing the first challenge. We tackle the second challenge by deriving rational agents adapted to these tasks using the framework of meta-learning, leading to a class of models called *ecologically rational meta-learned inference* (ERMI). ERMI quantitatively explains human data better than seven other cognitive models in two different experiments. It additionally matches human behavior on a qualitative level: (1) it finds the same tasks difficult that humans find difficult, (2) it becomes more reliant on an exemplar-based strategy for assigning categories with learning, and (3) it generalizes to unseen stimuli in a human-like way. Furthermore, we show that ERMI's ecologically valid priors allow it to achieve state-of-the-art performance on the OpenML-CC18 classification benchmark.

1. Introduction

Ecological rationality refers to the idea that humans are *rational* agents adapted to the *ecological* environments they interact with. Nearly seventy years ago, Brunswik (1955) emphasized that we have to move beyond laboratory settings and understand cognition in the light of naturalistic environments. Later on, Simon (1990) famously argued that human decision-making is like the two blades of a scissor, with one blade representing the cognitive processes of the mind and the other the structure of the environment in which the mind operates. Todd & Gigerenzer (2012) furthered this notion by introducing the term ecological rationality, suggesting that minds are adapted to their environments through the use of simple, context-specific strategies.

However, it has remained challenging to build computational models that describe how people implement strategies adapted to their environment for two reasons. First, defining ecologically valid tasks is still an open problem (Barker, 1968; Neisser, 1987; Hammond, 1998) and second, even if we have access to such tasks, it is challenging to build models that solve them rationally.

In the present paper, we address both of these challenges. We show that large language models (LLMs) – having been trained on large amounts of human-generated text – can serve as a useful tool for generating ecologically valid tasks, thereby addressing the first challenge. To address the second challenge, we then derive rational learning algorithms for these tasks using the framework of meta-learning (Pratt & Thrun, 1998; Hochreiter et al., 2001; Binz et al., 2023), leading to a class of models that we call *ecologically rational meta-learned inference* (ERMI).

We illustrate our approach using the domain of category learning (Ashby & Maddox, 2005) — one of the best-studied areas of cognitive science. We begin by verifying that LLMs can generate category learning tasks whose statistics match real-world classification data sets (Bischl et al., 2019). Following this, we show that ERMI quantitatively explains human data from two different category learning experiments better than seven other cognitive models. Furthermore, ERMI aligns with human behavior qualitatively: (1) it finds the same tasks difficult that humans find difficult, (2) it shows the same transition of categorization strategies as humans, and (3) it generalizes to unseen stimuli in a human-like way. Taken together, these results suggest that we can explain human category learning to a large extent using the principle of ecological rationality.

[*]Equal contribution [1]Computational Principles of Intelligence Lab, Max Planck Institute for Biological Cybernetics, Tübingen, Germany [2]Institute for Human-Centered AI, Helmholtz Computational Health Center, Munich, Germany. Correspondence to: Akshay K. Jagadish <akshaykjagadish@gmail.com>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

Furthermore, we hypothesized that the ecologically valid priors encoded in ERMI allow it to perform well on classification tasks from the machine learning literature. To test this hypothesis, we evaluate ERMI on the curated classification benchmark OpenML-CC18 (Bischl et al., 2019) and find that it achieves state-of-the-art performance.

2. Related work

LLMs for data generation: Recently, the wider concept of using LLM-generated data to train another model has become more popular (Gunasekar et al., 2023; Schick & Schütze, 2021; Wang et al., 2023; Bai et al., 2022; Mitra et al., 2023; Taori et al., 2023). For example, Gunasekar et al. (2023) prompted GPT-3.5 to generate synthetic textbook-quality data which they used to train a smaller transformer-based model. To justify this approach in the context of ecological rationality, one has to first establish that LLMs can produce ecologically valid tasks. Borisov et al. (2022) have done so recently by showing that LLMs are realistic tabular data generators, while Griffiths et al. (2023) demonstrated that LLM-generated data matches the priors of human subjects in several settings. (Coda-Forno et al., 2023) have shown that LLMs can even adapt their priors by meta-learning fully in-context.

Meta-learned models of cognition: Using models that achieve optimal task performance to study behavior is central to the rational analysis of cognition (Anderson, 1991b). Traditionally, these models have taken the form of Bayesian models (Griffiths et al., 2008). However, the Bayesian framework does not permit the construction of rational models for a given data set of tasks. The framework of meta-learning offers a way to overcome this problem (Binz et al., 2023). Unlike Bayesian models, meta-learned models of cognition can learn adaptive priors by repeatedly interacting with a distribution of tasks. Furthermore, these models have been shown to converge onto the optimal learning algorithm for the environments they are trained on (Ortega et al., 2019) and can be used in cases where the hand-crafting of assumptions is impractical or even infeasible.

Recently, it has been shown that meta-learned models capture human behavior across a wide range of domains, including decision-making (Binz et al., 2022a), reinforcement learning (Kumar et al., 2022; Binz & Schulz, 2022b; Jensen et al., 2023; Schubert et al., 2023), and compositional reasoning (Jagadish et al., 2023; Lake & Baroni, 2023). However, all these previous applications have relied on environments hand-engineered by researchers instead of ecologically valid ones.

Human category learning: How people learn to categorize objects has received significant attention in the cognitive sciences. For example, researchers have investigated how people learn to make fine-grained perceptual categorizations (Ashby & Townsend, 1986), what strategies people use when learning to categorize objects by comparing formal models of category learning (Smith & Minda, 1998; Nosofsky & Zaki, 2002; Maddox & Ashby, 1993), or whether there are different cognitive systems of category learning (Ashby & Maddox, 2005; Newell et al., 2011). In the present work, we make use of this rich literature by relying on its experimental paradigms and data. In particular, we used the paradigms developed by Shepard et al. (1961), Smith & Minda (1998), and Johansen & Palmeri (2002). We furthermore compare our model to a wide range of previously established category learning models (Nosofsky, 1986; Anderson, 1991a; Homa & Cultice, 1984; Nosofsky et al., 1994b).

3. Methods

In this section, we describe how we prompted LLMs to generate ecologically valid category learning tasks and how we then used meta-learning to learn models that are optimally adapted to these tasks.

3.1. Prompting LLMs to generate ecologically valid category learning tasks

A category learning task entails categorizing a stimulus $x \in \mathbb{R}^n$ into categories y based on its feature values. Multiple stimuli are presented sequentially and participants are tasked to predict their category after each presentation. Upon making their choice, they receive feedback on the true category of the stimulus and are presented with the next one.

To generate thousands of such category learning tasks from an LLM, we relied on a two-stage process. In the first stage, we queried the LLM to synthesize feature names and corresponding category labels. In the second stage, the model was prompted to produce data points for the feature names and category labels generated in the first stage.

More specifically, we used the following prompt to synthesize feature names and category labels:

Synthesize feature names and category labels

I am a psychologist who wants to run a category learning experiment. In a category learning experiment, there are many different three-dimensional stimuli, each of which belongs to one of two possible real-world categories.

Please generate names for three stimulus feature dimensions and two corresponding categories for 250 different category learning experiments:

The LLM then produced a series of category learning tasks

in a sequence until the specified number of tasks was generated. To illustrate one example, based on this prompt, the model constructed a category learning task with [SODIUM, FAT, PROTEIN] content as feature names and [HEALTHY, UNHEALTHY] as category labels.

In the second stage, we prompted the LLM to generate data points for a given category learning task:

Generate category learning tasks

I am a psychologist who wants to run a category learning experiment. For a category learning experiment, I need a list of stimuli and their category labels. Each stimulus is characterized by three distinct features: **sodium**, **fat**, and **protein**. These features can take only numerical values. The category label can take the values **healthy** or **unhealthy** and should be predictable from the feature values of the stimulus.

Please generate a list of 100 stimuli with their feature values and their corresponding category labels using the following template for each row:

– feature value 1, feature value 2, feature value 3, category label

Each generated data point contains feature values and their corresponding category label, e.g., [250, 15, 20, HEALTHY] for our previously mentioned example. In total, we generated three data sets containing around 10 000 different category learning tasks for three, four, and six feature dimensions. Each task consisted of 100, 300, and 616 data points, respectively. We provide further details about the generated category learning tasks (including the counts of the top 50 feature names and category labels) in Appendix A.

For our data generation procedure, we used CLAUDE-V2 (Anthropic, 2023) as it can process up to 100 000 tokens, is instruction-tuned, and performed well out of the box in our preliminary experiments. The temperature parameter was set to one to induce diversity and all other parameters were set to their default values. We provide details about other LLMs we considered and additional design choices in Appendix A.2.

3.2. Ecologically rational meta-learned inference (ERMI)

We parsed the generated tasks as described in Appendix A.3 and stored them in a numerical format. Then, we constructed rational learning algorithms for the numerical data by training memory-based meta-learning systems based on a two-stage process (Hochreiter et al., 2001; Santoro et al., 2016; Wang et al., 2016). In an inner-loop stage, a neural network predicts the category y_t for an input stimulus x_t conditioned on preceding stimulus-category pairs $x_{1:t-1}, y_{1:t-1}$. In an outer-loop stage, the network's parameters θ are updated using the following objective:

$$\arg\max_{\theta} \mathbb{E}_{p(x_{1:T}, y_{1:T})} \left[\sum_{t=1}^{T} \log p_\theta \left(y_t \mid x_{1:t}, y_{1:t-1} \right) \right] \quad (1)$$

where p_θ defines the output probabilities produced by the network.

During evaluation – i.e., once training is completed – the neural network implements a free-standing learning algorithm that can predict the category label of a new stimulus based on preceding stimulus-category pairs, despite its parameters being frozen. The resulting network approximates the Bayes-optimal learning algorithm for the data set of the category learning tasks $p(x_{1:T}, y_{1:T})$ encountered during training (Ortega et al., 2019).

We refer to the class of models derived by training on ecologically valid (i.e., LLM-generated) category learning task as *ecologically rational meta-learned inference* (ERMI). If trained on synthetically-generated category learning tasks sampled from a Bayesian logistic regression prior, we refer to the models as *meta-learned inference* (MI; Binz et al., 2022a). We chose MI as a baseline for two reasons. First, linear models have a long history as models of human learning (Karelaia & Hogarth, 2008; Lucas et al., 2015) and also category learning more specifically (Speekenbrink et al., 2008; 2010). Furthermore, the model has been previously used in a setting known as multiple cue probability learning (Binz et al., 2022a). While not exactly the same, this setting shares many similarities to category learning. Finally, using the terminology of Müller et al. (2022), we refer to the models as *prior-data fitted networks* (PFN) when tasks are sampled from a Bayesian neural network prior. For details on how these tasks are generated, see Appendix B.1.

The backbone for all our meta-learning models consisted of a transformer-based decoder architecture (Vaswani et al., 2017) with a causal attention mask. The network had six layers, a model dimension of 64, 256 hidden units in the feed-forward network, and eight attention heads. Positional encoding of input data points was done using sine and cosine functions of different frequencies (Vaswani et al., 2017). Note that during evaluation, transformer weights are frozen and learning is purely driven by self-attention applied to causally masked inputs.

In each training episode, a batch of tasks is sampled from $p(x_{1:T}, y_{1:T})$ and the model predicts the category for the given stimulus conditioned on all preceding stimulus-category pairs. Finally, the objective mentioned in Equation 1 is computed, and model parameters are updated using the ADAM optimizer (Kingma & Ba, 2014) with a learning rate

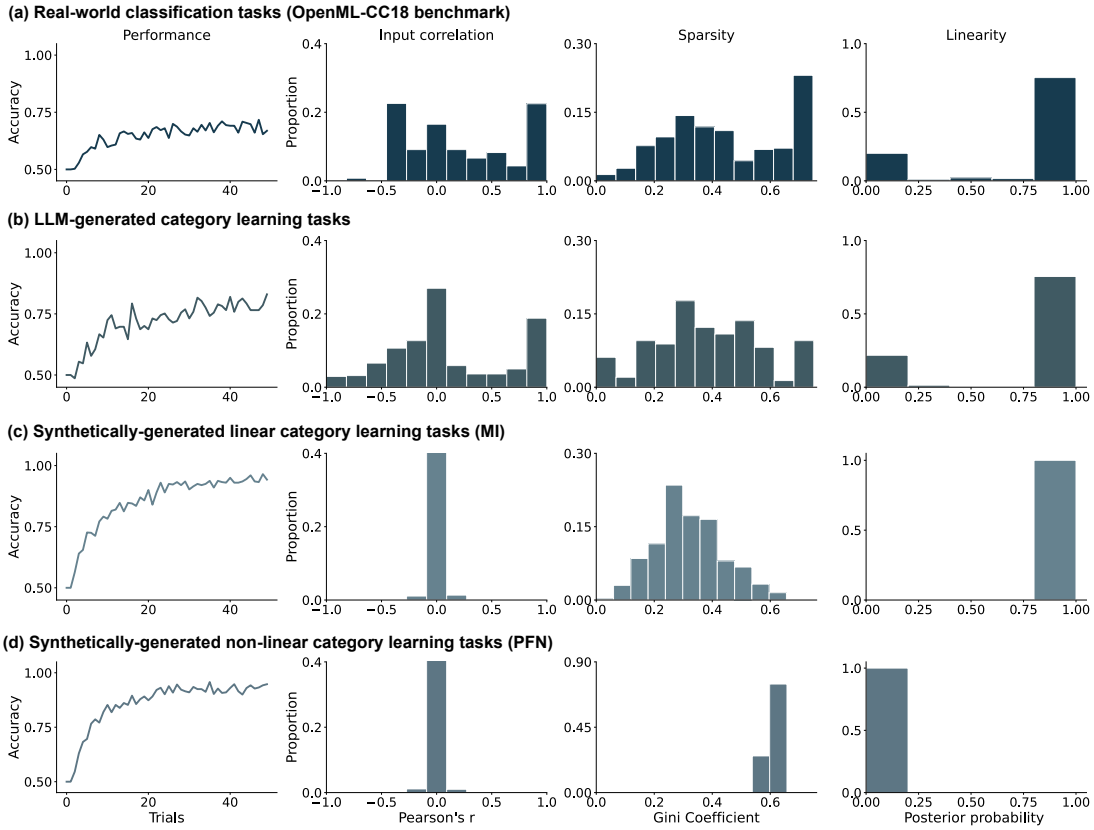

Figure 1. **LLM generates ecologically valid category learning tasks:** Mean task performance of the logistic regression model measured over trials (first column). Histogram of Pearson's correlation coefficients computed between pairs of features (second column). Histogram of Gini coefficients computed over the logistic regression weights (third column). Linearity of the category learning task (fourth column) computed for (a) 28 different real-world binary classification tasks from the OpenML-CC18 benchmarking suite (b) ecologically valid category learning tasks generated from CLAUDE-V2 and (c) synthetic category learning tasks derived using the Bayesian logistic regression prior that were used to train the meta-learned inference (MI) model (d) synthetic category learning tasks with nonlinear decision boundary derived using the Bayesian neural network prior that were used to train prior-fitted networks (PFN) model.

of 10^{-4}. This process is repeated for $500\,000$ episodes. We provide full details about the model training procedure in Appendix B.2.

4. LLM-generated category learning tasks are ecologically valid

ERMI can only be interpreted as an ecologically rational model if the statistics of LLM-generated tasks on which it was trained match the statistics of real-world classification problems. We verified that this was the case by comparing the data distributional properties of the two (Chan et al., 2022). For this analysis, we relied on the OpenML-CC18 benchmarking suite, a curated collection of real-world classification tasks (Bischl et al., 2019). Note that although the OpenML-CC18 benchmark is large enough to enable this form of statistical analysis, it is too small for direct meta-learning.

We downsampled all tasks in the OpenML-CC18 benchmark to four feature dimensions and included only binary classification tasks without any missing features in our analysis – amounting to 28 tasks. In addition, we also contrasted LLM-generated tasks to a collection of synthetically-generated category learning tasks with a linear decision boundary (cor-

responding to those used to train MI; refer to Appendix B.1). We analyzed these collections of tasks in terms of their learning curves, input correlations, sparsity, and linearity – details for which can be found in Appendix C.

Learning curves: Real data is noisy and not perfectly predictable. To investigate whether this is also true for our LLM-generated category learning tasks, we plotted the learning curves of a logistic regression model as a function of the number of training points. We found that the model reaches a ceiling accuracy of around 75% for both LLM-generated and real-world classification tasks (Figure 1; first column). In contrast, the ceiling performance for synthetically-generated tasks was much higher, reaching almost 100%.

Input correlations: Information contained in different feature dimensions are often correlated with each other. In the context of human cognition, it has been argued that this data distributional property of real-world data supports the reliance of people on heuristic decision-making strategies (Gigerenzer & Gaissmaier, 2011). While input dimensions in the synthetically-generated data were not correlated at all, both LLM-generated (0.11 ± 0.02; $t(1639) = 4.55$, $p < 0.001$) and real-world tasks (0.21 ± 0.01; $t(2252) = 14.64$, $p < 0.001$) showed a significant percentage of correlated features (Figure 1; second column). The corresponding histograms had similar shapes, both containing a peak at perfectly correlated features.

Sparsity: Another data distributional property that allows people to ignore information is sparsity — for many tasks only a few dimensions are relevant. We fitted a linear model on each task to evaluate whether we could find evidence for this in the LLM-generated data. We used the Gini coefficient – a measure borrowed from the economics literature – of the resulting regression coefficients to quantify sparsity (Binz et al., 2022a). High Gini coefficients correspond to maximal sparsity, meaning only a single feature is relevant. Both LLM-generated (0.38 ± 0.01; $t(545) = 3.81$, $p < 0.001$) and real-world tasks (0.45 ± 0.01; $t(762) = 10.83$, $p < 0.001$) exhibited significantly higher sparsity than synthetically-generated tasks (0.32 ± 0.01, see Figure 1; third column).

Linearity: People have strong priors towards linear relationships but can also learn non-linear ones given enough examples (Lucas et al., 2015; Brehmer, 1974). To measure whether this bias is also present in the distributional properties of the data, we conducted a model comparison between a linear model (a simple logistic regression model) and a non-linear one (logistic regression with higher-order polynomial features). For each task, we computed the posterior probability that the linear model offers a better explanation as our measure of linearity (details can be found in Appendix C). Most LLM-generated and real-world tasks were found to be linear but there was also a significant number of exceptions (Figure 1; fourth column). The synthetically-generated tasks, on the other hand, were fully linear by design.

Taken together, these analyses indicate that category learning tasks generated by LLMs share many features with real-world classification tasks. As they can also be produced in large quantities, these tasks can serve as a substitute for real-world classification tasks when meta-learning ecologically rational learning algorithms.

5. ERMI shows human-like learning difficulties

In the following sections, we investigate how well ERMI captures human category learning. We began by looking at one of the most canonical studies in category learning originally conducted by Shepard et al. (1961). The study required participants to learn how to categorize a stimulus that varied in shape (triangle or square), size (small or big), and color (black or white). In total, there were eight stimuli, and participants had to categorize them into one of two categories over several blocks of 16 trials. The authors assigned a stimulus to a category based on six different rules (labeled TYPE 1 to TYPE 6) that increased in difficulty. For more information, we refer to Appendix G.

Shepard et al. (1961) showed that people find tasks belonging to TYPE 1 the easiest and TYPE 6 the hardest, with the average error for TYPE 1 tasks going to zero after four blocks and TYPE 6 tasks remaining at around 10.6% even after 15 blocks (see Figure 2). The average error for TYPE 2 tasks (3.2%) was found to be lower than for TYPE 3 tasks (6.1%), TYPE 4 tasks (6.5%), and TYPE 5 tasks (7.5%).

We simulated ERMI and MI on the Shepard et al. (1961) study. Figure 2 (a) to (c) shows their learning curves alongside those of humans for the six difficulty levels. It can be seen that the learning curves of ERMI are difficulty-dependent and are in terms of mean-squared error (MSE) more similar to humans ($\text{MSE} = 0.03$) than MI ($\text{MSE} = 0.26$). Notably, ERMI, like humans, finds the TYPE 1 task easier than the TYPE 6 task and shows a similar clustering of learning curves for TYPE 2 to TYPE 5 tasks (see Table 5 for details). However, we also find that ERMI learns much faster than people: its performance plateaus after four blocks while humans continue learning until the end of the experiment for most types. Even though MI performs tasks of different difficulty levels to varying degrees of success, its learning curves do not match those of humans. For example, unlike humans, MI finds TYPE 2 tasks as difficult as TYPE 6 tasks and TYPE 4 tasks easier than TYPE 3 tasks.

Furthermore, we investigated how well ERMI explains human trial-by-trial choices on a quantitative level. For this, we considered human data from Badham et al. (2017) who

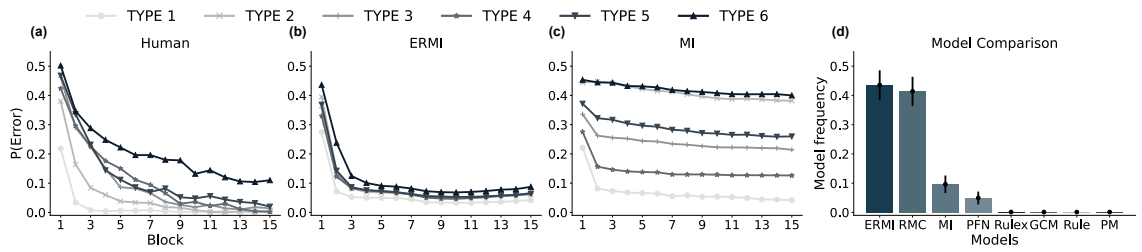

Figure 2. **ERMI shows human-like learning difficulties**: **(a-c)** Average error probabilities for each task TYPE in each block of 16 trials for (a) humans, (b) ERMI, and (c) MI. **(d)** The posterior model frequency of participants' choices in the Badham et al. (2017) study for eight computational models. Human data in (a) was reproduced from Table 1 in Nosofsky et al. (1994a). ERMI and MI were simulated on TYPE 1-6 tasks for 50 runs with the inverse temperature that resulted in the lowest mean-squared error compared to humans, which was $\beta = 0.4$ for ERMI, and $\beta = 0.9$ for MI.

conducted a replication of Shepard's original study that only included TYPE 1 to TYPE 4 tasks. We performed a Bayesian model comparison between eight computational models: the three meta-learned models introduced earlier (ERMI, MI, and PFN), and five other cognitive models. The five established category learning models from the cognitive science literature included the rational model of categorization (RMC; Anderson, 1991a), the generalized context model (GCM; Nosofsky, 1986), a prototype model (PM; Homa & Cultice, 1984), a rule-based model (Rule; Ashby & Townsend, 1986), and a rule-plus-exception model (Rulex; Nosofsky et al., 1994b). We provide more details about the five cognitive models, their fitting procedure, and the model comparison in Appendix D, E, and F.

We measured the goodness-of-fit to human choices based on two metrics: posterior model frequency and exceedance probability (Rigoux et al., 2014). The posterior model frequency measures how often a model offers the best explanation in the population, while the exceedance probability measures how likely it is that a given model is the most frequent explanation (the latter is reported in the Appendix G). Figure 2 (d) shows that ERMI explains human choices the best more frequently (0.43 ± 0.05) compared to the other models, with the RMC coming in a close second (0.41 ± 0.05). MI (0.10 ± 0.03) and PFN (0.05 ± 0.02) fit human data the best less than 10% of the time. The classical cognitive models like the exemplar-, prototype- and rule-based models failed at explaining human choices better than other competing models ($\leq 1\%$ of times).

Additionally, we simulated CLAUDE-V2 directly on the Badham et al. (2017) study (see Appendix G for details). We first observed that CLAUDE-V2 can perform the task optimally quite rapidly, with a mean error reaching zero for most difficulty levels already after three blocks. However, we also observed that the ordering of difficulty levels in CLAUDE-V2 does not match humans, with CLAUDE-V2 finding type 2 tasks more difficult than type 3, 4, and 5 tasks as shown in Figure 10. Furthermore, we found that the LLM's (MSE=0.18) match to human error rates was worse than that of ERMI (MSE=0.05), see Table 6 for details. We also fitted the CLAUDE-V2 predictions to human choices from the Badham et al. (2017) study and compared its performance to ERMI using the Bayesian information criterion (BIC; lower is better). We found that ERMI (38722.13) still offers a better fit to humans than CLAUDE-V2 (39816.90). Taken together, these results suggest that an LLM trained on the entire internet cannot explain human category learning as well as an ecologically rational meta-learning inference model.

6. ERMI becomes more exemplar-based with learning

What strategy people use to categorize objects and how the application of strategies changes over time are heavily debated questions in psychology. Smith & Minda (1998) attempted to understand whether people use prototype- or exemplar-based strategies during category learning. More specifically, they asked: do people learn a prototype for each category and assign categories based on the similarity of a stimulus to the learned prototypes, or do they instead remember previously seen examples for each category and assign categories based on the similarity of a stimulus to the stored exemplars?

To investigate this, Smith & Minda (1998) designed a category learning task that contained 14 six-dimensional stimuli, each of which was assigned to a category based on a nonlinear decision rule. Participants in their experiment were then tasked to assign one of the two categories to repeatedly presented stimuli. Following this, they fit predictions from prototype- and exemplar-based models to proportions of human choices, aggregated over trials within a block, by

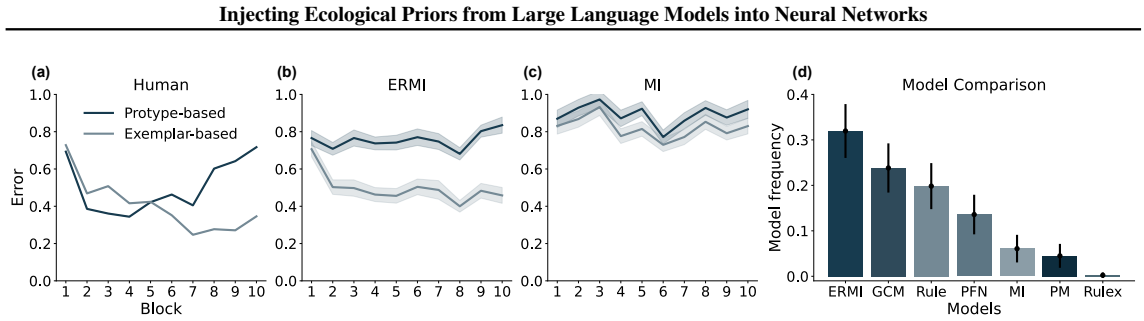

Figure 3. **ERMI becomes more exemplar-based with learning**: **(a-c)** The average error of exemplar- and prototype-based models fitted to (a) human choices, (b) simulated choices from ERMI, and (c) simulated choices from MI for each block of 56 trials. **(d)** The posterior model frequency of participants' choices in the Devraj et al. (2021) study for seven computational models. Human data in (a) was reproduced from Smith & Minda (1998). ERMI and MI were simulated using inverse temperature values fitted to participants' choices in Devraj et al. (2021). The mean of the fitted inverse temperature and its standard error were 0.09 ± 0.01 for ERMI and 0.17 ± 0.02 for MI, respectively. The shaded region shows the standard error of the mean.

minimizing the MSE between them. They found that people were better explained by the prototype-based model in the early blocks but in the later blocks, their choices aligned more closely with the exemplar-based model as shown in Figure 3 (a). We provide more details about this analysis in Appendix H.

We simulated choices from ERMI and MI on the task of Smith & Minda (1998) and fitted the prototype- and exemplar-based models on the simulated choices as in the original study. We find that ERMI, like humans, becomes increasingly exemplar-based over trial segments ($\hat{\beta} = -0.01 \pm 0.004; z = -2.54,\ p < 0.01$) whereas choices from MI are explained almost equally well by exemplar-based and prototype-based learning ($\hat{\beta} = -0.002 \pm 0.005; z = -0.47,\ p = 0.63$) as shown in Figure 3 (b) and (c). While humans are better explained by prototype-based models for the first five blocks, ERMI is already better explained by an exemplar-based model from the second block onwards. Like in the previous study, this again indicates that ERMI is learning the task faster than humans. Nonetheless, these results demonstrate that training on ecological category learning problems is sufficient for developing an exemplar-based strategy category assignment. We provide additional details and results in Appendix H.

We then evaluated if ERMI also explains human choices better than competing models in this study. To do this, we conducted a model comparison on human data from Devraj et al. (2021) – a replication of the Smith & Minda (1998) study – following the procedure outlined in Section 5. Posterior model frequency in Figure 3 (d) suggests that ERMI explains human choices the best most often (0.32 ± 0.06), closely followed by the GCM (0.24 ± 0.05) and the rule-based model (0.20 ± 0.05).

We additionally simulated CLAUDE-V2 directly on the Devraj et al. (2021) study. We found that CLAUDE-V2 is completely exemplar-based after the second block, which significantly differs from the strategy used by humans and ERMI (see Figure 12). We also found that ERMI (41207.20) offers a better fit to human choices than CLAUDE-V2 (42401.42) in terms of BIC. These results once again confirm that while CLAUDE-V2 can generate ecologically valid data, it cannot bring to bear its rich set of ecological priors to act in human-like ways.

7. ERMI displays human-like generalization

Having shown that ERMI learns category structures in a human-like way, we next inspect how it generalizes to stimuli unseen during training and whether it displays generalization patterns similar to people. To this end, we zoomed into the study from Johansen & Palmeri (2002), in which participants were instructed to categorize nine four-dimensional stimuli into two categories. The authors then examined how participants generalized to seven transfer stimuli (labeled T1-T7) for which they did not receive feedback during training. In Figure 4 (a), we report the mean probability of participants assigning category A to each of the seven transfer stimuli at the end of the experiment. It can be seen that they assigned stimuli T5, T6, and T7 mostly to category A and T1, T2, T3, and T4 mostly to category B. We provide further details about the paradigm in Appendix I.

We simulate the behavior of ERMI and MI on the Johansen & Palmeri (2002) study and found that ERMI generalizes to unseen stimuli in a human-like way by classifying stimuli T1, T3, and T4 more often as category B and T5, T6, and T7 more often as category A. MI, on the other hand, classified all stimuli except T7 mostly as category B. The Euclidean

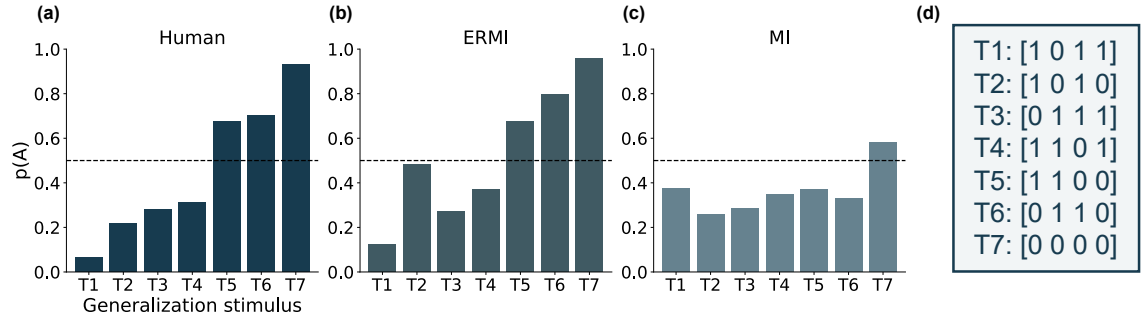

Figure 4. **ERMI displays human-like generalization: (a-c)** Average categorization probabilities of transfer stimuli T1-T7 for (a) humans (b) ERMI (c) MI. **(d)** The encoding scheme used for the seven transfer stimuli. Human data in (a) was reproduced from Johansen & Palmeri (2002). ERMI and MI were simulated on the same experiment for 77 runs, with inverse temperature settings that resulted in the lowest mean-squared error compared to humans, which was $\beta = 0.9$ for ERMI, and $\beta = 0.1$ for MI.

distance between the choice probabilities of humans and MI (0.67) was higher than that between humans and ERMI (0.29). The pattern of generalization of ERMI matches humans except for stimulus T2, which is classified at around chance level. Why exactly this is the case remains a question for future work. One possible explanation could relate to the observation that T2 only contains two non-zero features (see Figure 4 (d)) while all other stimuli categorized as B contain three non-zero features.

8. ERMI achieves state-of-the-art performance on machine learning benchmarks

Humans can bring to bear the rich set of priors they have acquired from their everyday interactions to generalize to novel tasks (Tenenbaum & Griffiths, 2001; Griffiths & Tenenbaum, 2006; Lake & Baroni, 2023). Thus far, we have demonstrated how ERMI captures some of these adaptive priors and explains essential aspects of human category learning. This led us to ask whether such a model can also perform well on real-world classification tasks from the machine learning literature.

To investigate this, we evaluated the performance of ERMI on a set of real-world classification tasks from the OpenML-CC18 benchmarking suite (Bischl et al., 2019). We excluded classification tasks with more than two classes, over 100 features, or missing values, resulting in a set of 23 classification tasks. We then compared the performance of ERMI against several baseline models, including logistic regression, a support vector machine (SVM; Cortes & Vapnik, 1995), XGBoost (Chen & Guestrin, 2016), and TabPFN (Hollmann et al., 2023). TabPFN is an off-the-shelf PFN-based model designed for tabular data prediction that has recently shown state-of-the-art performance on an independent large-scale evaluation (McElfresh et al., 2023).

Following the procedure of Müller et al. (2021), we created 20 class-balanced learning problems with 100 data points for each of the selected data sets. We provided 30 input-label pairs to our models for training and evaluated them on the remaining 70 data points. For each data set, we reduced the input dimensionality to four, keeping only the features with the highest F-value to the target variable. We measured the performance on the test set based on two metrics: accuracy and rank.

We found that ERMI is the best model in terms of both mean accuracy ($70.95\% \pm 0.54$) and mean rank (2.26 ± 0.22; see Table 7). TabPFN is the second-best model in terms of mean accuracy ($70.51\% \pm 0.63$), while XGBoost is the second-best in terms of mean rank (2.61 ± 0.30). We provide the summary of main results in Table 1 and detailed results across all data sets in Appendix J.

The performance gain of ERMI over TabPFN in terms of mean accuracy is in the same range as TabPFN over XGBoost, indicating that the improvement is substantial. In terms of parameters, TabPFN has 64 times more parameters than ERMI, and on disk, it is around 80 times larger, suggesting that there is room for further improvement. When compared to a parameter-matched PFN and MI, ERMI shows a significant accuracy boost of 3.5% and 10% respectively.

We additionally adapted a Bayesian analysis from (Stephan et al., 2009) to compute the probability that a model offers the best performance (within a set of alternative models) most frequently across data sets. This measurement is also known as the exceedance probability (EXP) in the statistics literature and is reported in the last row of Table 1. We found that ERMI is the best model according to this metric,

Table 1. Performance metrics on OpenML-CC18 benchmark.

MEAN	SVM	XGBOOST	TABPFN	ERMI
ACC.	69.29%	70.17%	70.51%	**70.95%**
RANK	2.76	2.61	2.85	**2.26**
EXP	0.01	0.00	0.24	**0.66**

with an exceedance probability of 0.66. TabPFN comes second with an exceedance probability of 0.24.

9. Discussion

Ecological rationality has a long history in cognitive science (Brunswik, 1955; Simon, 1990; Todd & Gigerenzer, 2012). The proposition that people are to some degree adapted to the problems they have to solve in the real world is maybe trivial. Figuring out how strong this adaptation is, on the other hand, has remained a big open question. Yet, it has been notoriously difficult to build models that are optimally adapted to the problems that people encounter in their everyday environment. From a technical perspective, the framework of meta-learning offers a solution to this problem. Yet, it has thus far only been applied to artificially-generated environments (Kumar et al., 2020; Binz & Schulz, 2022b; Binz et al., 2022a; Lake & Baroni, 2023; Jagadish et al., 2023). The main obstacle up to now was that the number of available real-world data sets was insufficient for meta-learning. We have shown that one can overcome this obstacle by prompting LLMs to generate a large collection of ecologically valid category learning tasks. We have then used meta-learning to obtain models that are optimally adapted to them, leading to a class of ecologically rational models that we call ERMI.

ERMI captured three patterns, which are also observed when humans learn to categorize objects: (1) it showed similar learning difficulties as humans, (2) it became more exemplar-based as learning progressed, and (3) it displayed human-like generalization patterns. Furthermore, it explained human behavior better than competing approaches on a quantitative level, thereby suggesting that we can explain many characteristics of human category learning using the principle of ecological rationality.

The methodology developed, more importantly, enables us to test whether people are ecologically rational or not, thereby allowing us to ask questions such as: how much of human learning can be attributed to data distributional properties alone? The approach we have proposed is quite general and in future work, we plan to extend it to other domains, such as decision-making (Bourgin et al., 2019; Peterson et al., 2021), reinforcement learning (Brändle et al., 2021), and function learning (Schulz et al., 2017; 2016). Furthermore, it would be interesting to not impose a predefined task structure on the LLM but instead let it synthesize arbitrary task structures by itself.

While it is possible to come up with a mathematical formulation that allows generating features that match a given statistic like correlation or linearity, it is unclear how to design an algorithm that generates category-learning tasks that match all real-world statistics measured and unmeasured. This is exactly where the strength of pre-trained LLMs lies. If queried correctly, LLMs can act as a single, easy-to-access, rich – possibly infinite – source of real-world data.

10. Limitations

There are some limitations to generating ecological data that are worth noting. Firstly, it is uncertain whether the current approach can be scaled to generate several more orders of features and multi-class category learning problems. Secondly, strict data curation protocols are necessary if the approach is to be used in cases where unbiased factually correct data are required. Thirdly, preliminary analyses revealed that the data generated by LLMs are US/western-centric (Anthropic, 2023; Tamkin et al., 2023). Therefore, culturally diverse LLMs are needed to generate data that capture the rich diversity of the real world.

On the cognitive science front, there were also facets of human category learning not captured by ERMI. In particular, we found that ERMI generally learned faster than people. In the third experiment, for instance, ERMI already displayed the generalization patterns shown in Figure 4 after two blocks, while people required 16 blocks or more (Johansen & Palmeri, 2002). We believe that this gap can (at least partially) be closed by incorporating limited computational resources (Binz et al., 2022a; Jagadish et al., 2023) or other architectural constraints (Achterberg et al., 2023).

11. Conclusion

We have shown that LLMs can generate ecologically valid category learning tasks that can be used for meta-learning. With these models at hand, we then demonstrated that one can explain human category learning to a large extent using the principle of ecological rationality. Furthermore, the priors acquired by ERMI are rich enough that it achieves state-of-the-art performance on a real-world classification benchmark. In future work, we plan to scale up ERMI's architecture, train it on classification tasks with a flexible number of features, increase the maximum number of data points, and allow for more than two classes.

Acknowledgements

We thank all reviewers for their constructive and thoughtful feedback. We would also like to thank the authors of Devraj et al. (2021), Nosofsky et al. (1994a), and Badham et al. (2017) for making the data from their study available. Furthermore, we thank the members of the "Computational Principles of Intelligence Laboratory" (CPI Lab), participants of the "Analytical Connectionism" summer school, Dan John, Yashas Annadani, and Laura Heidiri for their comments, discussions, and support. This work was supported by the Max Planck Society, the Volkswagen Foundation, and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645.15/18.

Impact Statement

This paper presents work whose goal is to show that methods from machine learning, specifically large language models and meta-learning, can be used to advance our understanding of human cognition. An important finding from our work that might be of significance to a broader audience is that training on a large corpus of human-generated text has allowed large language models to capture certain key ecological features. Therefore, one has to be aware that such very powerful features can emerge in foundation models when scaled appropriately. There may be other potential societal consequences of our work, but we do not feel that they need to be specifically highlighted here.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Achterberg, J., Akarca, D., Assem, M., Heimbach, M., Astle, D. E., and Duncan, J. Building artificial neural circuits for domain-general cognition: a primer on brain-inspired systems-level architecture. *arXiv preprint arXiv:2303.13651*, 2023.

Anderson, J. R. The adaptive nature of human categorization. *Psychol. Rev.*, 98(3):409–429, July 1991a. ISSN 0033-295X, 1939-1471. doi: 10.1037/0033-295X.98.3.409.

Anderson, J. R. Is human cognition adaptive? *Behavioral and brain sciences*, 14(3):471–485, 1991b.

Anthropic, P. B. C. Claude 2. https://www.anthropic.com/index/claude-2, July 2023. Accessed: 2024-1-15.

Ashby, F. G. and Maddox, W. T. Human Category Learning. *Annual Review of Psychology*, 56(1):149–178, February 2005. ISSN 0066-4308, 1545-2085. doi: 10.1146/annurev.psych.56.091103.070217. URL https://www.annualreviews.org/doi/10.1146/annurev.psych.56.091103.070217.

Ashby, F. G. and Townsend, J. T. Varieties of perceptual independence. *Psychological Review*, 93(2):154–179, 1986. ISSN 1939-1471. doi: 10.1037/0033-295X.93.2.154. Place: US Publisher: American Psychological Association.

Badham, S. P., Sanborn, A. N., and Maylor, E. A. Deficits in category learning in older adults: Rule-based versus clustering accounts. *Psychol. Aging*, 32(5):473–488, August 2017. ISSN 0882-7974, 1939-1498. doi: 10.1037/pag0000183.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.

Barker, R. G. *Ecological psychology: Concepts and Methods for Studying the Environment of Human Behavior.* Stanford, CA: Stanford University Press, 1968.

Binz, M. and Schulz, E. Modeling human exploration through resource-rational reinforcement learning. *Advances in Neural Information Processing Systems*, 35:31755–31768, 2022b.

Binz, M., Gershman, S. J., Schulz, E., and Endres, D. Heuristics from bounded meta-learned inference. *Psychological review*, 2022a.

Binz, M., Dasgupta, I., Jagadish, A., Botvinick, M., Wang, J. X., and Schulz, E. Meta-learned models of cognition. *arXiv preprint arXiv:2304.06729*, 2023.

Bischl, B., Casalicchio, G., Feurer, M., Hutter, F., Lang, M., Mantovani, R. G., van Rijn, J. N., and Vanschoren, J. Openml benchmarking suites. *arXiv:1708.03731v2 [stat.ML]*, 2019.

Borisov, V., Seßler, K., Leemann, T., Pawelczyk, M., and Kasneci, G. Language models are realistic tabular data generators. *arXiv preprint arXiv:2210.06280*, 2022.

Bourgin, D. D., Peterson, J. C., Reichman, D., Russell, S. J., and Griffiths, T. L. Cognitive model priors for predicting human decisions. In *International conference on machine learning*, pp. 5133–5141. PMLR, 2019.

Brändle, F., Binz, M., and Schulz, E. Exploration beyond bandits. *The drive for knowledge: The science of human information seeking*, pp. 147–168, 2021.

Brehmer, B. Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance*, 11 (1):1–27, 1974.

Brunswik, E. Representative design and probabilistic theory in a functional psychology. *Psychological review*, 62(3): 193, 1955.

Chan, S., Santoro, A., Lampinen, A., Wang, J., Singh, A., Richemond, P., McClelland, J., and Hill, F. Data distributional properties drive emergent in-context learning in transformers. *Advances in Neural Information Processing Systems*, 35:18878–18891, 2022.

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

Coda-Forno, J., Binz, M., Akata, Z., Botvinick, M., Wang, J. X., and Schulz, E. Meta-in-context learning in large language models. *arXiv preprint arXiv:2305.12907*, 2023.

Cortes, C. and Vapnik, V. Support-vector networks. *Machine learning*, 20:273–297, 1995.

Daunizeau, J., Adam, V., and Rigoux, L. Vba: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS computational biology*, 10(1): e1003441, 2014.

Devraj, A., Zhang, Q., and Griffiths, T. The dynamics of exemplar and prototype representations depend on environmental statistics. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43, 2021.

Gigerenzer, G. and Gaissmaier, W. Heuristic decision making. *Annual review of psychology*, 62:451–482, 2011.

Griffiths, T., Kemp, C., and B Tenenbaum, J. *Bayesian models of cognition*. Carnegie Mellon University, 2008.

Griffiths, T. L. and Tenenbaum, J. B. Optimal predictions in everyday cognition. *Psychological science*, 17(9):767–773, 2006.

Griffiths, T. L., Zhu, J.-Q., Grant, E., and McCoy, R. T. Bayes in the age of intelligent machines. *arXiv preprint arXiv:2311.10206*, 2023.

Gunasekar, S., Zhang, Y., Aneja, J., Mendes, C. C. T., Del Giorno, A., Gopi, S., Javaheripi, M., Kauffmann, P., de Rosa, G., Saarikivi, O., et al. Textbooks are all you need. *arXiv preprint arXiv:2306.11644*, 2023.

Hammond, K. R. Ecological validity: Then and now, 1998.

Hochreiter, S., Younger, A. S., and Conwell, P. R. Learning to learn using gradient descent. In *Artificial Neural Networks—ICANN 2001: International Conference Vienna, Austria, August 21–25, 2001 Proceedings 11*, pp. 87–94. Springer, 2001.

Hollmann, N., Müller, S., Eggensperger, K., and Hutter, F. TabPFN: A transformer that solves small tabular classification problems in a second. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=cp5PvcI6w8_.

Homa, D. and Cultice, J. C. Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1):83, 1984.

Jagadish, A. K., Binz, M., Saanum, T., Wang, J. X., and Schulz, E. Zero-shot compositional reinforcement learning in humans. *PsyArXiv preprint PsyArXiv:ymve5*, 2023.

Jensen, K. T., Hennequin, G., and Mattar, M. G. A recurrent network model of planning explains hippocampal replay and human behavior. *bioRxiv*, pp. 2023–01, 2023.

Johansen, M. K. and Palmeri, T. J. Are there representational shifts during category learning? *Cogn. Psychol.*, 45 (4):482–553, December 2002. ISSN 0010-0285. doi: 10.1016/s0010-0285(02)00505-4.

Karelaia, N. and Hogarth, R. M. Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological bulletin*, 134(3):404, 2008.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kumar, S., Dasgupta, I., Cohen, J., Daw, N., and Griffiths, T. Meta-learning of structured task distributions in humans and machines. In *International Conference on Learning Representations*, 2020.

Kumar, S., Correa, C. G., Dasgupta, I., Marjieh, R., Hu, M. Y., Hawkins, R., Cohen, J. D., Narasimhan, K., Griffiths, T., et al. Using natural language and program abstractions to instill human inductive biases in machines. *Advances in Neural Information Processing Systems*, 35: 167–180, 2022.

Lake, B. M. and Baroni, M. Human-like systematic generalization through a meta-learning neural network. *Nature*, pp. 1–7, 2023.

Lucas, C. G., Griffiths, T. L., Williams, J. J., and Kalish, M. L. A rational model of function learning. *Psychonomic bulletin & review*, 22(5):1193–1215, 2015.

Maddox, W. T. and Ashby, F. G. Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, 53(1):49–70, January 1993. ISSN 0031-5117, 1532-5962. doi: 10.3758/BF03211715. URL http://link.springer.com/10.3758/BF03211715.

McElfresh, D., Khandagale, S., Valverde, J., Ramakrishnan, G., Goldblum, M., White, C., et al. When do neural nets outperform boosted trees on tabular data? *arXiv preprint arXiv:2305.02997*, 2023.

Medin, D. L. and Schaffer, M. M. Context theory of classification learning. *Psychol. Rev.*, 85(3):207–238, May 1978. ISSN 0033-295X, 1939-1471. doi: 10.1037/0033-295x.85.3.207.

Mitra, A., Del Corro, L., Mahajan, S., Codas, A., Simoes Ribeiro, C., Agrawal, S., Chen, X., Razdaibiedina, A., Jones, E., Aggarwal, K., Palangi, H., Zheng, G., Rosset, C., Khanpour, H., and Awadallah, A. Orca-2: Teaching small language models how to reason. arXiv, November 2023.

Müller, S., Hollmann, N., Arango, S. P., Grabocka, J., and Hutter, F. Transformers can do bayesian inference. *arXiv preprint arXiv:2112.10510*, 2021.

Müller, S., Hollmann, N., Arango, S. P., Grabocka, J., and Hutter, F. Transformers can do bayesian inference. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=KSugKcbNf9.

Neisser, U. *Cognition and reality*. W H Freeman/Times Books/ Henry Holt and Co, 1987.

Newell, B. R., Dunn, J. C., and Kalish, M. Chapter six - Systems of Category Learning: Fact or Fantasy? In Ross, B. H. (ed.), *Psychology of Learning and Motivation*, volume 54 of *Advances in Research and Theory*, pp. 167–215. Academic Press, January 2011. doi: 10.1016/B978-0-12-385527-5.00006-1. URL https://www.sciencedirect.com/science/article/pii/B9780123855275000061.

Nosofsky, R. M. Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, 115(1):39, 1986.

Nosofsky, R. M. and Zaki, S. R. Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(5):924, August 2002. ISSN 1939-1285. doi: 10.1037/0278-7393.28.5.924. URL https://psycnet.apa.org/fulltext/2002-15432-008.pdf. Publisher: US: American Psychological Association.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., and Glauthier, P. Comparing models of rule-based classification learning: a replication and extension of shepard, hovland, and jenkins (1961). *Mem. Cognit.*, 22 (3):352–369, May 1994a. ISSN 0090-502X. doi: 10.3758/bf03200862.

Nosofsky, R. M., Palmeri, T. J., and McKinley, S. C. Rule-plus-exception model of classification learning. *Psychol. Rev.*, 101(1):53–79, January 1994b. ISSN 0033-295X. doi: 10.1037/0033-295x.101.1.53.

Ortega, P. A., Wang, J. X., Rowland, M., Genewein, T., Kurth-Nelson, Z., Pascanu, R., Heess, N., Veness, J., Pritzel, A., Sprechmann, P., et al. Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030*, 2019.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D., and Griffiths, T. L. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science*, 372(6547):1209–1214, 2021.

Pratt, L. and Thrun, S. *Learning to learn*. Kluwer Academic Publishers, 1998.

Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. Bayesian model selection for group studies—revisited. *Neuroimage*, 84:971–985, 2014.

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pp. 1842–1850. PMLR, 2016.

Schick, T. and Schütze, H. Generating datasets with pretrained language models. *arXiv preprint arXiv:2104.07540*, 2021.

Schubert, J., Jagadish, A., Binz, M., and Schulz, E. A rational analysis of the optimism bias using meta-reinforcement learning. In *Conference on Cognitive Computational Neuroscience (CCN 2023)*, 2023.

Schulz, E., Tenenbaum, J., Duvenaud, D. K., Speekenbrink, M., and Gershman, S. J. Probing the compositionality of intuitive functions. *Advances in neural information processing systems*, 29, 2016.

Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., and Gershman, S. J. Compositional inductive biases in function learning. *Cognitive psychology*, 99:44–79, 2017.

Shepard, R. N., Hovland, C. I., and Jenkins, H. M. Learning and memorization of classifications. *Psychological Monographs: General and Applied*, 75(13):1–42, 1961. ISSN 0096-9753. doi: 10.1037/h0093825.

Simon, H. A. Invariants of human behavior. *Annual review of psychology*, 41(1):1–20, 1990.

Smith, J. D. and Minda, J. P. Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, memory, and cognition*, 24(6): 1411, 1998.

Speekenbrink, M., Channon, S., and Shanks, D. R. Learning strategies in amnesia. *Neuroscience & Biobehavioral Reviews*, 32(2):292–310, 2008.

Speekenbrink, M., Lagnado, D. A., Wilkinson, L., Jahanshahi, M., and Shanks, D. R. Models of probabilistic category learning in parkinson's disease: Strategy use and the effects of l-dopa. *Journal of Mathematical Psychology*, 54(1):123–136, 2010.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. Bayesian model selection for group studies. *Neuroimage*, 46(4):1004–1017, 2009.

Tamkin, A., Askell, A., Lovitt, L., Durmus, E., Joseph, N., Kravec, S., Nguyen, K., Kaplan, J., and Ganguli, D. Evaluating and mitigating discrimination in language model decisions. *arXiv preprint arXiv:2312.03689*, 2023.

Taori, R., Gulrajani, I., Zhang, T., Dubois, Y., Li, X., Guestrin, C., Liang, P., and Hashimoto, T. B. Stanford alpaca: An instruction-following llama model. https://github.com/tatsu-lab/stanford_alpaca, 2023.

Tenenbaum, J. B. and Griffiths, T. L. Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, 24(4):629–640, 2001.

Todd, P. M. and Gigerenzer, G. *Ecological rationality: Intelligence in the world*. Oxford University Press, Cary, NC, February 2012. ISBN 9786613594013.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Wang, J. X., Kurth-Nelson, Z., Tirumala, D., Soyer, H., Leibo, J. Z., Munos, R., Blundell, C., Kumaran, D., and Botvinick, M. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*, 2016.

Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., and Wei, F. Improving text embeddings with large language models. *arXiv preprint arXiv:2401.00368*, 2023.

A. Generating ecologically valid category learning tasks using LLMs

A.1. Synthesizing task features and labels

We synthesized task features and labels from CLAUDE-V2 using the prompt mentioned in Section 3.1, running it for a total of 100 batches. In each batch, we generate 250 tasks or until a maximum token length of 10k is reached. We repeat the procedure for all three different stimuli dimensions. In total, we synthesized 23421, 20690, and 13693 category learning tasks with three, four, and six-dimensional features respectively.

We show the counts for the top-50 most frequently occurring task features in Figure 5 and categories in Figure 6 for the 23421, 20690, and 13693 category learning tasks generated with three (a), four (b), and six-dimensional features respectively. We found that the model tends to produce features belonging to topics such as musicality (for instance, rhythm, melody, lyrics, tempo, vocals) and food (for instance, aroma, texture, crust, diet, protein). Regarding categories, there were many related to music (for example, classical, pop, jazz, rock) and vehicles (like trucks, SUVs, sedans). In future work, we plan to do a semantic analysis of the generated task features and category labels using methods such as hierarchical clustering to study the semantic grouping of the generated task features/categories.

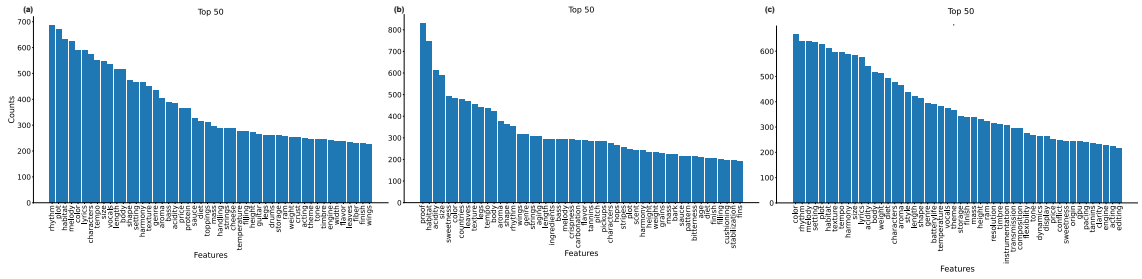

Figure 5. **Frequency of different features in CLAUDE-V2 synthesized category learning tasks:** Counts for the top-50 most frequently occurring task features in the 23421, 20690, and 13693 category learning tasks generated for three (a), four (b), and six-dimensional features respectively.

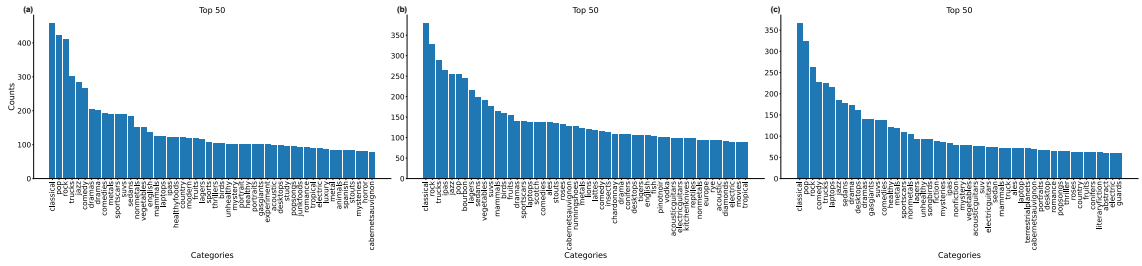

Figure 6. **Frequency of different categories in CLAUDE-V2 synthesized category learning tasks:** Counts for the top-50 most frequently occurring category labels in the 23421, 20690, and 13693 category learning tasks generated for three (a), four (b), and six-dimensional features respectively.

A.2. Generating category learning tasks

We used the prompt mentioned in Section 3.1 to generate data points for task features and labels synthesized in the first stage. We aimed to generate 100, 300, and 616 data points for category learning tasks with stimuli of three, four, and six dimensions respectively. However, sometimes the upper bound of 100k tokens is reached and the model does not generate the number of data points specified in the prompt. This was especially the case in category learning tasks with higher dimensional stimuli. In these cases, we generate the data points in two steps. In step one, we do the data generation as

before using the original prompt. In step two, we query the model again but now conditioned it on the first 20-40 percent of the data points generated in step 1 along with the prompt. This way, we could scale up the generated data points to up around 1.5 times the original length keeping the same underlying data distribution. We generated a total of 11518, 8950, and 12911 category learning tasks with three, four, and six-dimensional stimuli respectively.

Note on other LLMs: We ran preliminary tests on LLMs other than CLAUDE-V2 for generating category learning tasks including LLaMA (Touvron et al., 2023) and GPT-4 (Achiam et al., 2023). The non-instruction-tuned LLaMA version was not able to consistently produce the 100-616 data points we need per task. It was also especially difficult to parse the output of the model as the generated output failed to stick to the provided format. This problem could be mitigated with instruction-tuned versions of LLaMA that are now available but we leave it for future work. We also performed preliminary tests on the GPT-4 model from OpenAI for category learning task generation but found the model sampled the values for the features from uniform distribution using the code module and applied simple heuristic rules most of the time. For example, the sum of two features should be greater than the third and the mean of the two features greater than the other. Upon conducting a preliminary statistical analysis on a relatively small data set generated from GPT-4, we found that task statistics are similar to the statistics of the MI data set, thereby lacking the diversity in terms of measures reported in Section 4.

A.3. Parsing data generated by LLMs

Parsing synthesized task features and labels: We queried the CLAUDE-V2 to synthesize task features and labels in the following format: FEATURE DIMENSION 1, FEATURE DIMENSION 2, ..., FEATURE DIMENSION N, CATEGORY LABEL 1, CATEGORY LABEL 2. We then used a regular expression (regex) pattern, specifically `\d+\. (.+?)\n`, to efficiently parse and extract relevant data from the model output. This regex was designed to identify and isolate sequences beginning with a number followed by a period, capturing subsequent characters up to the first newline character. The extracted text was then processed to acquire the names for feature dimensions and category labels by splitting the string at the commas. The final processed data is then stored as a dataframe for future use.

Parsing generated task data points: We queried the CLAUDE-V2 to generate data points for a given category learning task in the following format: - FEATURE VALUE 1, FEATURE VALUE 2,..., FEATURE VALUE N, CATEGORY LABEL. The model followed the aforementioned format while generating the data points for a category learning task, more often than not. We then used a collection of regular expression (regex) patterns to parse the generated, ensuring accurate handling of different data formats. These regex patterns were designed to cover a wide range of scenarios: capturing numeric values with or without decimal points, handling alphanumeric strings including those with hyphens, and handling complex cases involving commas, hyphens, and optional preceding labels. Furthermore, they extend to different delimiters and formats, from simple comma-separated values to more complex structures with optional components. In Table 2, we show all the regex expressions used to parse data points. Based on these expressions, we were able to successfully parse up to 95 % of the tasks generated by the model. The values extracted from these regex expressions are then stored in a dataframe which acts as an offline repository of tasks on which one can train the ecologically rational meta-learned inference model.

B. Meta-learned inference models

B.1. Synthetic data generation

Bayesian logistic regression prior used for training MI model: We generated 10k synthetic binary classification tasks with a linear decision boundary using a Bayesian logistic regression model. To do this, we sample the input features from a normal distribution with zero mean and unit variance for a given number of data points and stimulus dimensions. We then applied a linear transformation, followed by a sigmoid function, and rounded the result to determine the binary class for the given input. The parameters of the linear transformation are sampled from a normal distribution with zero mean and unit variance. The maximum number of data points within a task was set to 400, 650, or 300 for category learning tasks with three, four, and six-dimensional stimuli respectively. These values were chosen depending on the length of the experiments on which these models were evaluated.

Bayesian neural network prior used for training PFN model: We generated 10k synthetic binary classification tasks using a version of the Bayesian neural network (BNN) prior developed by Müller et al. (2021). We used normally-distributed i.i.d. input features for a given number of data points and stimulus dimensions. We then pass the input through a BNN with two layers with tanh non-linearity and hidden dimensionality of 64. The network weights and biases were sampled from a

Table 2. Regular expression patterns used for parsing the data points generated for category learning tasks by CLAUDE-V2

INDEX	REGULAR EXPRESSION
1	`([\d.]+),([\d.]+),([\d.]+),([\w]+)`
2	`([\w\-]+),([\w\-]+),([\w\-]+),([\w]+)`
3	`([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+)`
4	`([^,]+),([^,]+),([^,]+),([^,]+)`
5	`([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+)`
6	`(?:.*?:)?([^,-]+),([^,-]+),([^,-]+),([^,-]+)`
7	`([^,-]+),([^,-]+),([^,-]+),([^,-]+)`
8	`r'^(\d+):([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\w]+)'`
9	`r'^(\d+):([\w\-]+),([\w\-]+),([\w\-]+),([\w\-]+),([\w]+)'`
10	`r'^(\d+):([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+)'`
11	`r'^(\d+):([^,]+),([^,]+),([^,]+),([^,]+),([^,]+)'`
12	`r'^(\d+):([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+)'`
13	`r'^(\d+):(?:.*?:)?([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)'`
14	`r'^(\d+):([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)'`
15	`^(\d+):([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\w]+)`
16	`^(\d+):([\w-]+),([\w-]+),([\w-]+),([\w-]+),([\w-]+),([\w-]+),([\w]+)`
17	`(\d+):([^,]+),([^,]+),([^,]+),([^,]+),([^,]+),([^,]+),([^,]+)`
18	`(\d+):([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+)`
19	`(\d+):(?:.*?:)?([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)`
20	`(\d+):([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)`

normal distribution with a mean of zero and standard deviation of 0.1 and subjected to an additional sparsity constraint (i.e., 20 percent of randomly chosen network weights and biases set to zero). The maximum number of data points was once again set to 400, 650, or 300 for category learning tasks with three, four, and six-dimensional stimuli respectively. The model output is passed through a sigmoid function to generate probability estimates which are then rounded to determine the class for the given input.

B.2. Data pre-processing, model architecture, and training

Data pre-processing: We filter out all tasks with more than two unique category labels and then binarize the category labels which are originally strings to make them consistent across tasks. The assignment of category labels, that is either '0' or '1', within a category learning task was randomized during batch creation. This ensures that there can be no unintended correlations between the stimuli seen during training and the labels (across all training data each input vector is assigned half of the time to label '0' and half of the time to label '1'). We also normalized each feature independently using a min-max normalization scheme such that values taken by any feature lie always between zero and one. Both the task features and data points were shuffled while generating tasks. Note that the tasks generated by LLMs are typically of different lengths. Whenever the sampled tasks are of variable lengths, they are padded with zeros to match the length of the longest task sample within the batch. We additionally also sampled LLM-generated data points with replacement to match the length of the experimental task used in the Devraj et al. (2021) and Johansen & Palmeri (2002) studies. We resorted to this strategy as the LLM-generated tasks had a maximum of about 200 data points per task and by resampling, we can evaluate the model on experiments with larger horizons without any drop in performance. The batch size was set to 64 for three- and four-dimensional stimuli and to 32 for six-dimensional stimuli and it operated under a maximum steps regime of 400, 300, and 650 for three, four, and six-dimensional tasks respectively.

Model architecture and training: The task features were mapped onto a 64-dimensional embedding space and positional encoded using sine and cosine functions of different frequencies as in Vaswani et al. (2017). A causal attention mask was then generated for the inputs such that the model makes conditional predictions on all preceding data points. The inputs along the attention mask are then passed to the transformer decoder model which had six layers, a model dimension of 64, 256 hidden units in the feed-forward network, and eight attention heads. The output of the transformer was then passed through a linear readout and sigmoid function to generate probability estimates for category 1. In practice, inference for all time steps is performed in parallel by passing a causal attention mask to the TRANSFORMERDECODER module in PyTorch (Paszke et al., 2019). We used binary cross-entropy (BCE) loss for a given batch of inputs and updated the model parameters using the ADAM optimizer (Kingma & Ba, 2014) with a learning rate of 10^{-4}. We trained all our models for a total of

500 000 episodes.

C. LLMs generate ecologically valid category learning tasks

Sparsity: We fitted a logistic regression model for each task and analyzed the sparsity of the resulting regression weights $\mathbf{w} \in \mathbb{R}^d$ using the Gini coefficient G:

$$G(\mathbf{w}) = \frac{\sum_{i=1}^{d}\sum_{j=1}^{d}|\mathbf{w}_i - \mathbf{w}_j|}{2d\sum_{i=1}^{d}\mathbf{w}_i} \tag{2}$$

Linearity: We fitted a logistic regression model and a logistic regression with second-order polynomial features on data $\mathcal{D}$ from each task. We then computed the Bayesian information criterion (BIC) for both models and used them to approximate the posterior probability that the linear model offers a better explanation of the data (assuming a uniform prior over models):

$$p(M = \text{linear}|\mathcal{D}) \approx \frac{\exp(-0.5 \cdot \text{BIC}_{\text{linear}})}{\sum_{m\in\{\text{linear, polynomial}\}}\exp(-0.5 \cdot \text{BIC}_m)} \tag{3}$$

Features values: We compared the distributions of input values generated by LLMs to input values from real-world datasets and found them to be extremely similar, as shown in Figure 7. We furthermore checked if there were spurious values in the LLM-generated data such as peaks at multiples of 5s and 10s, and found them to be at chance level (20% for multiples of 5s, 10% for multiples of 10s).

Additional analysis: We furthermore computed the KL divergence between the histograms of the real-world data set and the synthetically generated data sets for each investigated metric. The resulting KL divergences confirm that the LLM-generated data provides a good proxy for real-world data (see Table 3). In Figure 8, we overlaid the histograms of the real-world data set and the synthetically generated data sets for the four different statistics discussed in the main paper.

Table 3. KL divergence between real-world data classification data and the synthetically generated data for different statistical measures

REAL WORLD TASKS VS.	INPUT CORRELATION	GINI COEFFICIENT	POSTERIOR PROBABILITY	FEATURES
LLM-GENERATED	0.2716	0.2286	0.0202	0.0418
MI PRIOR	7.7637	2.8170	2.5265	1.6758
PFN PRIOR	7.7746	8.6391	2.5265	1.6775

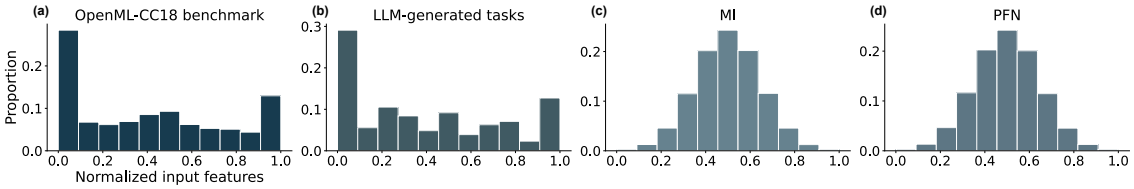

Figure 7. **Comparing input features of real-world classification tasks and LLM-generated tasks**: (a) Histogram of normalized input features from 28 different real-world binary classification tasks from the OpenML-CC18 benchmarking suite, (b) from ecologically valid category learning tasks generated from CLAUDE-V2, (c) synthetic category learning tasks derived using the Bayesian logistic regression prior that were used to train the meta-learned inference (MI) model and (d) synthetic category learning tasks with nonlinear decision boundary derived using the Bayesian neural network prior that were used to train prior-fitted networks (PFN) model.

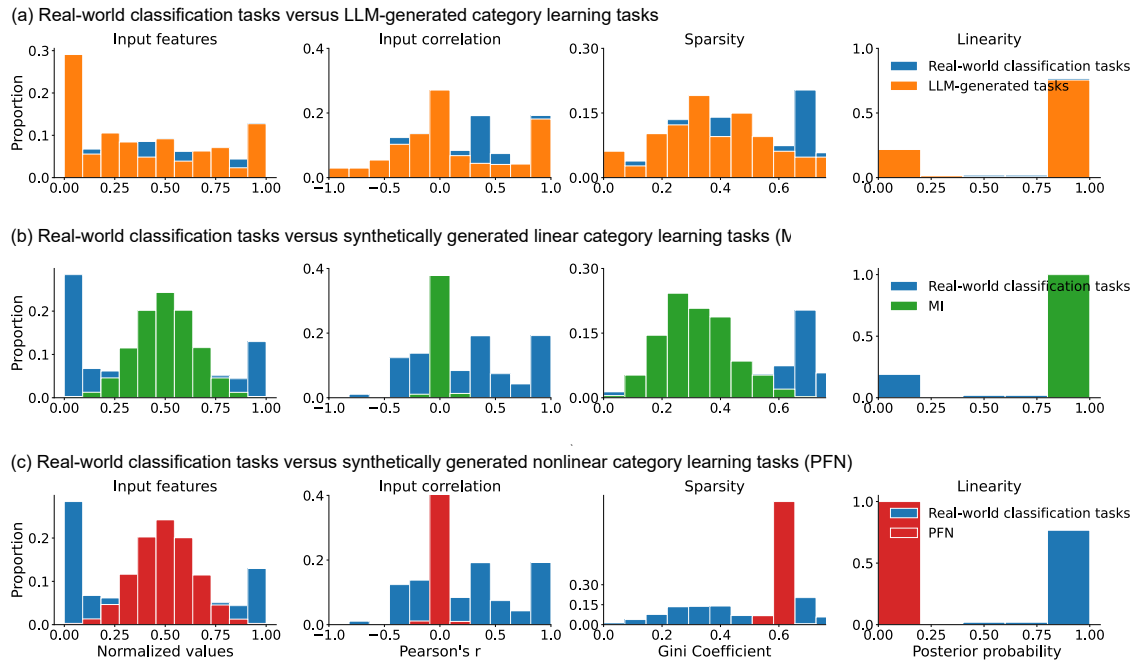

Figure 8. **Comparing the ecological validity of different category learning tasks:** Histogram of normalized input features (first column). Histogram of Pearson's correlation coefficients computed between pairs of features (second column). Histogram of Gini coefficients computed over the logistic regression weights (third column). Linearity of the category learning task (fourth column). These four statistics were (a) ecologically valid category learning tasks generated from CLAUDE-V2, (b) synthetic category learning tasks derived using the Bayesian logistic regression prior that were used to train the meta-learned inference (MI) model, and (c) synthetic category learning tasks with nonlinear decision boundary derived using the Bayesian neural network prior that were used to train prior-fitted networks (PFN) model were overlaid on the same statistics computed on 28 different real-world binary classification tasks from the OpenML-CC18 benchmarking suite.

D. Cognitive models

In this section, we will provide details regarding the five cognitive models we used for model comparison.

Rational model of categorization (RMC): The RMC is a Bayesian model of human category learning (Anderson, 1991a). In this paper, we used a meta-learned version of the model, which was obtained using the following data-generating distribution described in Badham et al. (2017). Model architecture and training followed the protocol used for ERMI, MI, and PFN. We set the free parameters based on an earlier study (Nosofsky et al., 1994a) to the following values: $c = 0.318$, $s_P = 0.488$, and $s_L = 0.046$. Note that we did not account for these parameters in our model comparisons, which slightly overestimates the ability of the RMC to explain human behavior.

Prototype-based model (PM): There are several versions of the prototype model (Medin & Schaffer, 1978; Smith & Minda, 1998). Here, we use the version from Smith & Minda (1998). The prototype model assigns a category to an observed stimulus based on the similarity to the category prototypes. The raw distance between the stimulus and a prototype, q_k, for category k is computed as a weighted sum of absolute differences across n feature dimensions with weights, $w_j \in [0, 1]$, for the features contained to sum to 1 as shown in Equation 4.

$$d_{x,q_k} = \sum_{j=1}^{n} w_j \left| x_j - q_{k,j} \right|, \tag{4}$$

Note that prototypes themselves can either be learned or provided during model definition. Here, we learn the prototypes for the two categories $\{q_1, q_2\}$. Therefore, $q_{k,j} \in [0, 1.]\ \forall j = \{1, 2, ...n\}$ are themselves also parameters. Distance is then converted to psychological similarity between prototypes and stimuli using:

$$\eta_{x,q_k} = e^{-c \cdot d_{x,q_k}} \tag{5}$$

where c is a sensitivity parameter that can shrink or amplify discriminability in psychological space. The probability of the stimulus being assigned to category 1 is then computed using:

$$P(k = 1 \mid x) = \frac{\eta_{x,q_1}}{\eta_{x,q_1} + \eta_{x,q_2}} \tag{6}$$

Furthermore, the final model predicted likelihood is a mixture between the predicted probability from the model and a random guess, with guessing parameter ϵ controlling the mixture probabilities:

$$p(k = 1 \mid x) = (1 - \epsilon) P(k = 1 \mid x) + \epsilon \cdot \mathrm{K}^{-1} \tag{7}$$

where K indicates number of categories.

Generalized context model (GCM): We used the GCM developed by Nosofsky (1986). The GCM categorizes an observed stimulus to a category by comparing the sum of its similarity to all previously seen exemplars in each category, $\{C_1, C_2\}$. The raw distance and similarity between observed stimulus and exemplars were computed based on Equations 4 and 5 respectively. The probability of assigning a stimulus to category $k = 1$ is then computed based on the summed category similarities in the following way:

$$P(k = 1 \mid x) = \frac{\sum_{y \in C_1} \eta_{x,y}}{\sum_{y \in C_1} \eta_{x,y} + \sum_{y \in C_2} \eta_{x,y}} \tag{8}$$

The final model predicted category probabilities is again a mixture between the predicted probability from the model and a random guess as mentioned in Equation 7.

Rule (Rule): The rule model implemented in this work assigns a category based on one of the two rules, whichever explains the participant data better. The first rule is based on the values taken by stimulus features along one dimension, and the second is based on the application of the conjunctive rule on pairs of features – whether a given pair of stimulus features take on the same value. The final model predicted category probabilities is again a mixture between the category prediction from the model and a random guess as mentioned in Equation 7.

Rule plus exception model (Rulex): We use the same implementation as the Rule model but provide exceptions as inputs to the model. For Devraj et al. (2021) task, we provide $[1, 1, 1, 1, 0, 1]$ as the exception stimulus for category 1, and for category 2, it was set to $[0, 0, 0, 1, 0, 0]$. For Badham et al. (2017) task, we provide $[1, 1, 1]$ and $[0, 0, 0]$ as exceptions for TYPE-2 task. The final model predicted category probabilities is again a mixture between the category prediction from the model and a random guess as mentioned in Equation 7. In future work, we plan to implement a more detailed version of rule-plus-exception model from (Nosofsky et al., 1994b) where the model learns the exceptions along with the rule.

CLAUDE-V2: We used CLAUDE-V2 as a cognitive model of human category learning. To do that, we prompted CLAUDE-V2 by paraphrasing the instructions given to participants, see Appendix G and H for task-specific prompts. In essence, the model was instructed via the prompt that it would be presented with stimuli belonging to two categories and would receive feedback regarding the correct category following each stimulus. It was instructed to learn a rule based on stimulus features to assign the stimulus to the correct category with increasing experience. As the Claude API only returns the CLAUDE-V2 output token and not its probability, the model's prediction was coded as a binary variable, $\pi(k = 1 \mid x_t; x_{1:t-1}, y_{1:t-1})$. The final model predicted category probabilities is again a mixture between the category prediction from the model and a random guess as mentioned in Equation 9.

$$p(k = 1 \mid x_t) = (1 - \epsilon)\pi(k = 1 \mid x_t; x_{1:t-1}, y_{1:t-1}) + \epsilon \cdot \mathrm{K}^{-1} \tag{9}$$

E. Fitting models to human data

In this section, we explain the fitting procedure used to fit the parameters of the models to human data. The model parameters were fit to the data using maximum likelihood estimation. We explain the implementation details for the different model classes below. The full list of fitted parameters for each model is shown in Table 4.

MI, PFN, RMC and ERMI: We fit an inverse temperature term β within the sigmoid function, which squashes the output from the final layer of the transformer to be within $[0, 1]$, to each participant. Note that this term is set to one during the meta-learning phase to derive an optimal model and is fitted only during the evaluation phase with the rest of the model weights being frozen. We use the differential evolution optimizer available in the SciPy optimization library (Virtanen et al., 2020) for fitting.

GCM and PM: Both models predict the probability of picking a category in a trial-by-trial fashion conditioned on all preceding stimulus-target pairs. We fit their parameters to human choices that minimize the negative log-likelihood of human choices under the model prediction. To do so, we used the MINIMIZE module available in SciPy's optimization library. As mentioned in Section D, the weights for the features were bounded to be within $[0, 1]$ and sum to 1, the sensitivity term bounded to be within $[0, 20]$. The prototype model requires learning the prototypical stimulus for each category (same dimensionality as the input stimulus, with the feature values bounded to be within $[0, 1]$). Both models learning a guessing parameter ϵ, which was bounded to be within $[0, 1]$.

Rule and Rulex: We used the same procedure as above except that we learn the stimulus dimension v_i on which the rule is applied.

CLAUDE-V2: We used the same procedure as above except that only the guessing parameter, ϵ, is learned.

F. Bayesian model comparison

In this section, we provide details regarding the Bayesian model comparison procedure used to compare the fits of different models to the behavioral data. We first performed maximum likelihood estimation to fit model parameters θ_m. We then computed the Bayesian information criterion (BIC) for model m for a given participant as follows:

$$\mathrm{BIC}_m = -2 \cdot \max_{\theta_m} \sum_{t=1}^{T} \log p_{\theta_m}\left(\hat{y}_t \mid x_{1:t}, y_{1:t-1}\right) + |\theta_m| \log(T) \tag{10}$$

where $|\theta_m|$ is the number of parameters estimated for model m, T is the number of trials in the task and $\hat{y}_t$ is the category

Table 4. Fitted parameters in each model where β is the inverse temperature term, w_i indicates the weights for the stimulus feature dimension i, n is the number of stimulus feature dimensions, c is the sensitivity term, ϵ is noise term in an epsilon greedy policy, q_1 and q_2 are the values for the prototypes for d stimulus features, and v_i are the stimulus dimension on which the rule is applied.

MODEL	PARAMETERS	
ERMI, MI, PFN, RMC	β	
GCM	c, ϵ, w_i	$\forall\, i \in \{1, 2, \ldots, n\}$
PM	$c, \epsilon, w_i, q_{1,i}, q_{2,i}$	$\forall\, i \in \{1, 2, \ldots, n\}$
RULE	v_1, v_2, ϵ	
RULEX	v_1, v_2, ϵ	
CLAUDE-V2	ϵ	

choice made by the participant in trial t. BIC penalizes the model based on its complexity and can be used as a measure for comparing goodness-of-fit when models differ in terms of their number of parameters.

We reported two metrics in the paper: posterior model frequency (in the main text) and exceedance probability. To compute them, we used a Python implementation of the Variational Bayesian Analysis (VBA) toolbox (Daunizeau et al., 2014). The toolbox requires us to provide log-evidences for each model and participant, which we approximate using $-0.5 \cdot \text{BIC}_m$. For further details about this model comparison procedure, see Rigoux et al. (2014).

G. ERMI shows human-like difficulty effects

G.1. Experiment details for Shepard et al. (1961) and Nosofsky et al. (1994a)

In their replication of the Shepard et al. (1961) study, Nosofsky et al. (1994a) conducted the study on 120 participants. The authors used geometric stimuli that varied in shape (squares or triangles), interior line type (solid or dotted), and size (large or small). Every participant completed two problems, therefore, each problem was performed by 40 participants. The participants were informed that the rules for each problem were independent. Following the methodology of Shepard et al. (1961), the learning process involved classifying stimuli into two categories and receiving feedback. This process was repeated over several blocks (containing up to 16 trials) with randomized stimulus order in each block. Learning in the task was measured until participants achieved a no-error streak in four consecutive sub-blocks of eight trials or reached a maximum of 400 trials. For more details, please refer to Nosofsky et al. (1994a).

In tasks belonging to TYPE 1, stimuli were assigned to a category depending on the values they take along one of the three dimensions, whereas in TYPE 2 tasks, stimuli were assigned to a category by applying the exclusive-or rule along two relevant dimensions. Category assignment in tasks belonging to TYPE 3, TYPE 4, and TYPE 5 used a unidimensional rule-plus-exception structure with some stimuli grouped in the central region and some in the periphery. Lastly, TYPE 6 tasks require considering feature values along all dimensions. For the illustration of category structures for the six types, please refer to Figure 1 in (Nosofsky et al., 1994a).

In Badham et al. (2017), the authors replicated the Shepard et al. (1961) study on 96 adults aged between 18 to 87 years. They used eight geometric shapes varying in size (large or small), shape (square or triangle), and color (black or white) in the experiment with the stimuli shown on a mid-gray background. The order of stimuli and their category assignment were randomized. They only considered the first four types from the Shepard et al. (1961) study but unlike their study, participants performed all four types. Participants performed each task type for a total of six blocks with each block containing 16 trials (resulting in a total of 96 trials) or until they reached a criterion of perfect performance in two consecutive blocks. For more details, please refer to Badham et al. (2017).

G.2. Simulations:

To run simulations of the Shepard et al. (1961) study on ERMI, MI, and PFN model, the geometric stimuli used in the experiment as mentioned above are converted into binary coded vectors taking values along the three stimulus feature dimensions. The value assignment for a stimulus feature was randomized in every run, the order of presentation of the stimulus was also randomized, and the number of presentations of a stimulus per block was matched to the original study. In each run, the model was evaluated on a task of one particular type.

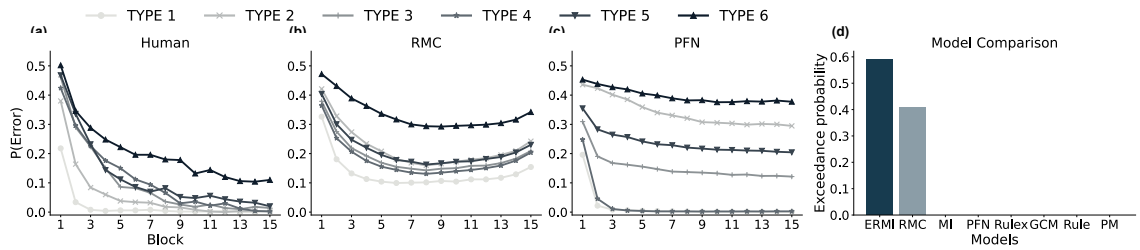

Figure 9. **Supplementary figure accompanying Figure 2**: **(a-c)** Average error probabilities for each task TYPE in each block of 16 trials for (a) humans, (b) RMC, and (c) PFN. **(d)** The exceedance probability of participants' choices in the Badham et al. (2017) study for eight computational models. Human data in (a) was reproduced from Table 1 in Nosofsky et al. (1994a). RMC and PFN were simulated on TYPE 1-6 tasks for 50 runs with the inverse temperature that resulted in the lowest mean-squared error compared to humans, which was $\beta = 0.9$ for ERMI, and $\beta = 0.9$ for MI.

Table 5. Mean performance of humans and models for each rule type in replication of Shepard et al. (1961) study over 15 blocks. Human data was taken from Table 1 in Nosofsky et al. (1994a). Details of model simulations can be found in Appendix G.

Model	**Rule**						**MSE**
	TYPE 1	TYPE 2	TYPE 3	TYPE 4	TYPE 5	TYPE 6	
Humans	.0201	.0565	.1015	.1120	.1212	.2048	.0000
ERMI	.0586	.0891	.0855	.0826	.0888	.1172	**.0287**
MI	.0686	.4089	.2404	.1431	.2880	.4201	.2627
PFN	.0170	.3405	.1533	.0226	.2371	.3975	.1736
RMC	.1329	.2215	.1903	.1718	.2132	.3364	.1003

G.3. Additional observations and results

Why are TYPE 2 and TYPE 6 hard? We think that this could be because TYPE 2 tasks involve applying the exclusive-or rule along two relevant dimensions (and ignoring one of the dimensions altogether), while TYPE 6 tasks require memorizing feature values taken by stimuli along all dimensions, making it hard for models to learn.

Learning curves of RMC and PFN are not as similar to humans as ERMI: MSE between learning curves of humans and RMC and PFN was 0.10 and 0.17 respectively. They are larger than that of ERMI which was 0.03.

ERMI learns the task faster than people: We transform the block variable t using an exponential kernel as follows:

$$y = ae^{-b*t} + c \tag{11}$$

where a is the amplitude, b is the decay coefficient term, and c is the offset term, and then regressed the transformed variable onto the error rate for both ERMI and humans. We found that the fitted decay coefficient for ERMI (1.24) is larger than for humans (0.44).

Baseline models fit the data adequately: Baseline models particularly the GCM (177.35 ± 4.70) and PM (201.19 ± 3.38) model did fit the data quite well in terms of log-likelihoods compared to ERMI (200.82 ± 4.56). However, the number of parameters being fit to human data is quite large in these models (refer to Table 4). Therefore, they are heavily penalized and have higher BIC values than ERMI.

G.4. CLAUDE-V2 as a cognitive model of human category learning

Simulations: To run simulations of the Badham et al. (2017) study on CLAUDE-V2, we queried it using the prompt shown in Appendix G.4. The geometric stimuli used in the experiment are described in textual format. The order of stimulus

presentation was randomized, and the number of stimulus presentations per block was matched to the original study. We ran 96 simulation runs for each of the six rules.

Prompt for Badham et al. 2017 study

In this experiment, you will be shown examples of geometric objects. Each object has three different features: size, color, and shape. Your job is to learn a rule based on the object features that allows you to tell whether each example belongs in the {A} or {B} category. As you are shown each example, you will be asked to make a category judgment and then you will receive feedback. At first you will have to guess, but you will gain experience as you go along. Try your best to gain mastery of the {A} and {B} categories.

- In trial 1, you picked category {A} for Big Black Square and category {A} was correct.
- In trial 2, you picked category {A} for Small Black Triangle and category {B} was correct

Human: What category would a Small Black Triangle belong to? (Give the answer in the form "Category ⟨your answer⟩").
Assistant: Category

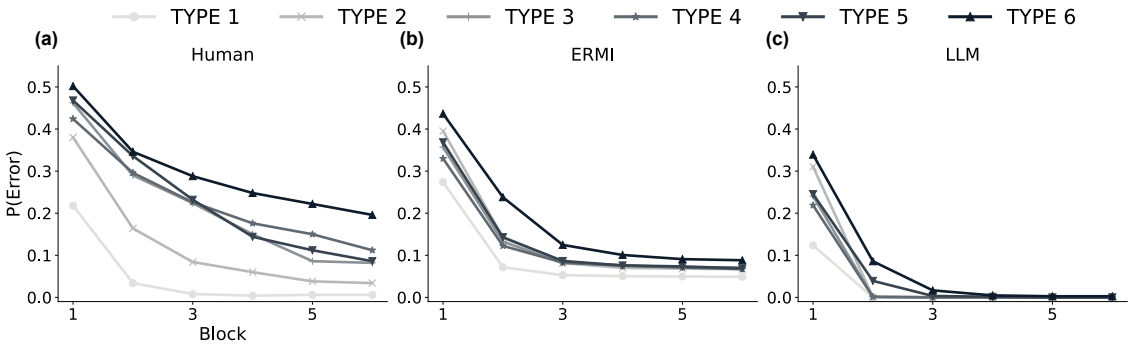

Figure 10. **Unlike ERMI, CLAUDE-V2 does not show human-like learning difficulties**: **(a-c)** Average error probabilities for each task TYPE in each block of 16 trials for (a) humans, (b) ERMI, and (c) LLM. Human data in (a) was reproduced from Table 1 in Nosofsky et al. (1994a). ERMI was simulated on TYPE 1-6 tasks for 50 runs with the inverse temperature set to $\beta = 0.4$. CLAUDE-V2 was simulated for 94 runs each on TYPE 1-6 tasks with temperature term set to 0.

Table 6. Mean performance of humans and models for each rule type in the Badham et al. (2017) study. Details of model simulations can be found in Appendix G.

Model	**Rule**						**MSE**
	TYPE 1	TYPE 2	TYPE 3	TYPE 4	TYPE 5	TYPE 6	
Human	.0460	.1267	.2157	.2307	.2297	.3003	.0000
ERMI	.0912	.1374	.1317	.1248	.1361	.1798	**.0538**
LLM	.0206	.0521	.0404	.0368	.0482	.0752	.1771
MI	.0959	.4337	.2655	.1657	.3175	.4387	.1795
PFN	.0402	.3908	.1887	.0527	.2722	.4237	.1541
RMC	.1593	.2744	.2314	.2164	.2576	.3845	.0564

H. ERMI becomes more exemplar-based

H.1. Experiment details for Smith & Minda (1998) and Devraj et al. (2021)

Smith & Minda (1998) conducted their study on 32 participants, using 14 six-dimensional stimuli, with each stimulus mapping to a six-letter nonsensical word such as gafuzi, kafitdo, nivety, wysero, etc. — see Appendix A of Smith & Minda (1998) for all words. Each stimulus can be represented by a six-digit binary string, where each digit and position corresponds to a specific letter. For example, if the stimulus 'gafuzi' corresponds to the binary code '000000', then 'gyfuzi' corresponds to '010000', and so on. The stimuli were assigned to categories such that stimulus '000000' corresponds to category 1 and stimulus '11111' corresponds to category 2. We only considered the non-linearly separable (NLS) category structure from Experiment 2 in this work. According to this, a category contained five stimuli with five features in common with the prototype, and one stimulus with five features in common with the opposing prototype. Therefore, category 1 contained seven stimuli as follows: [000000, 100000, 010000, 001000, 000010, 000001, 111101]. The remaining seven stimuli belonged to category 2 [111111, 011111, 101111, 110111, 111011, 111110, 000100]. Participants had unlimited time to make their choice on each trial. After making their choice, they were told whether it was a correct decision or not. Participants completed a total of 560 trials, or 10 blocks (called trial segments by the authors but we call them blocks to be consistent with other experiments) of 56 trials each, in which they saw each stimulus four times.

Devraj et al. (2021) replicated the task as mentioned above and collected data from 60 participants. Participants were recruited from the 18-23 age range and English-speaking population using Prolific. Their study involved 11 blocks instead of 10 and as a result, they had 616 trials.

H.2. Model-based analysis:

The 616 choices made by participants and meta-learning models were divided into 11 blocks of 56 trials each. We obtained the choices from the models – ERMI, MI, and PFN – by simulating them on the task for a total of 50 runs using the βs fitted to participants in the Devraj et al. (2021) study. We then fit prototype- and exemplar-based models onto the choices of humans and models to see if they are better explained by prototype or exemplar-based strategy. To fit their parameters, we minimize the sum of squared errors (SSE) between observed and predicted probabilities for each participant for a given block following the original study's approach:

$$SSE = \sum_{t=1}^{14} (p(k=1|x_t) - \hat{p}_{1,x_t})^2 \tag{12}$$

where $p(k=1|x_t)$, from Equation 7, is the predicted probability from the model – either GCM or PM – that stimulus x_t belongs to category 1 based on an entire trial segment (56 trials) of data, and $\hat{p}_{1,x_t}$ is the proportion of trials in the trial segment (out of those in which stimulus x_t was seen) in which the participant or model categorized stimulus x_t to category 1. We used SciPy's Sequential Least Squares Programming (SLSQP) method in the SciPy's optimization module to obtain the best fitting parameter for the two models as in (Devraj et al., 2021). We then compared the SSE computed using the best-fitting parameters between the two models as shown in Figure 3(a).

H.3. Additional observations and results

ERMI is better explained by the exemplar-based model with learning: We fit a linear mixed-effects model to measure this effect quantitatively. We predict the average error term from the GCM and PM model using blocks and model as predictors and test the interaction between blocks and model. Blocks (1-10) were mean-centered and the exemplar model (GCM) was coded as -1 and the prototype-based model (PM) was coded as 1. We report the coefficient $\hat{\beta}$ for the interaction term in the main paper. We found that ERMI ($\hat{\beta} = -0.01 \pm 0.004; z = -2.54,\ p < 0.01$) is better explained by the exemplar-based model with learning whereas choices from MI are explained equally well by exemplar-based and prototype-based learning ($\hat{\beta} = -0.002 \pm 0.005; z = -0.47,\ p = 0.63$) as shown in Figure 3

Prototype model cannot learn exceptions: Given that the category structure is non-linearly separable (containing exceptions), a prototype model cannot explain the data fully even if provided the true category choices as it tends to miscategorize the exception stimulus from which category. The exemplar-based model (GCM) model, however, has no such issues and can fit the true choices perfectly.

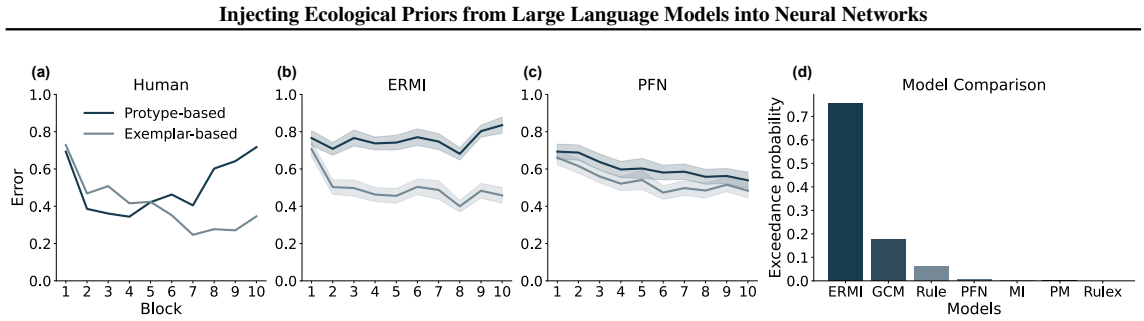

Figure 11. **Supplementary material accompanying Figure 3:(a-c)** The average error of exemplar- and prototype-based models fitted to (a) human choices, (b) simulated choices from ERMI, and (c) simulated choices from PFN for each block of 56 trials. **(d)** The exceedance probability of participants' choices in the Devraj et al. (2021) study for seven computational models. Human data in (a) was reproduced from Smith & Minda (1998). ERMI and MI were simulated using inverse temperature values fitted to participants' choices in Devraj et al. (2021). The mean of the fitted inverse temperature and its standard error were 0.09 ± 0.01 for ERMI and 0.14 ± 0.01 for MI, respectively. The shaded region shows the standard error of the mean.

MI and PFN models find it hard to learn exceptions: A better fit of the prototype model to the MI and PFN in the latter half of the experiment (as shown in Figure 3 and 11) suggests that, as observed in Shepard's task, they are not able to learn exceptions like the prototype model as mentioned before.

H.4. CLAUDE-V2 as a cognitive model of human category learning

Simulations: To run simulations of the Smith & Minda (1998) study on CLAUDE-V2, we queried it using the prompt shown in Appendix H.4. The nonsense stimuli used in the experiment are provided in text. The order of presentation of the stimulus was also randomized, and the number of stimulus presentations in a block was matched to the original study. We ran the model for a total of 120 simulation runs.

Prompt for Devraj et al. 2022 study

In this experiment, you will be shown examples of nonsense word stimuli. Look carefully at each word and decide if it belongs to group {U} or group {M}. Respond with {U} if you think it is a group {U} word and {M} if you think it is a group {M} word. You will recieve feedback about the correct group after each of your response. At first, the task will seem quite difficult, but with time and practice, you should be able to answer correctly.

- In trial 1, you picked group {M} for wafuzi and group {U} was correct.
- In trial 2, you picked group {M} for gyfuzi and group {U} was correct.

Human: What group would the word gyfuzi belong to? (Give the answer in the form "Group ⟨your answer⟩").
Assistant: Group

I. ERMI shows human-like generalization

I.1. Experiment details for Johansen & Palmeri (2002)

Johansen & Palmeri (2002) conducted their study with 198 participants (out of which 68 were excluded for further analysis) using four-dimensional stimuli with each dimension taking one of two possible values. The stimuli were "computer-generated drawings of rockets that varied along four binary-valued dimensions: The shape of the wing (triangular or rectangular), tail (jagged or boxed), nose (staircase or half-circle), and porthole (circular or star)" (Johansen & Palmeri, 2002). The category structure used in the study was similar to the ones used in classical studies such as Medin & Schaffer (1978); Nosofsky et al. (1994b) and is ill-defined in that no single feature along a dimension can be used to perfectly classify stimuli. Rather, the categories have a family resemblance structure in that category A stimuli tend to have a value of 0 along each dimension, and category B stimuli tend to have a value of 1 along each dimension. In this case, category 1 contained five stimuli as

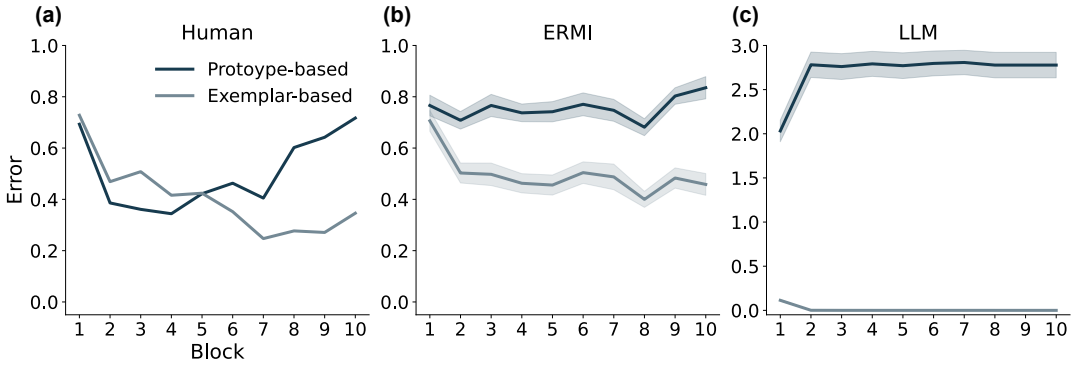

Figure 12. **CLAUDE-V2 performs purely exemplar-based category learning**: **(a-c)** The average error of exemplar- and prototype-based models fitted to (a) human choices, (b) simulated choices from ERMI, and (c) simulated choices from CLAUDE-V2 for each block of 56 trials. Human data in (a) was reproduced from Smith & Minda (1998). ERMI was simulated using inverse temperature values fitted to participants' choices in Devraj et al. (2021). The mean of the fitted inverse temperature and its standard error for ERMI was 0.09 ± 0.01. CLAUDE-V2 was queried using the prompt shown in Appendix H.4 with the temperature term set to 0 for 120 runs. The shaded region shows the standard error of the mean.

follows: [0001, 0101, 0100, 0010, 1000]. The remaining four stimuli belonged to category 2 [0011, 1001, 1110, 1111]. The stimulus presentation order was randomized within each block. Participants had unlimited time to make their choice on each trial. After making their choice, they were told whether or not it was a correct choice. Participants completed a total of 288 training trials, or 32 blocks of 9 trials each, in which they saw each stimulus once. However, in addition to the training block, participants had to perform a transfer block after 2, 4, 8, 16, 24, and 32 blocks of training. In a transfer block, all 16 possible stimuli are shown without any corrective feedback.

I.2. Simulations

We simulated ERMI, MI, and PFN on the Johansen & Palmeri (2002) study for different betas values, from zero to one in steps of 0.1, for a total of 544 runs. The models interacted with each of the nine training stimuli 32 times with the ordering of the stimuli shuffled between runs. Predictions for the transfer stimuli were derived by concatenating them – one at a time – at the end of 32 training blocks in every run. By doing so, we were able to derive the model's prediction for each unseen stimulus around 77 times. In Figure 4 and 13, we reported average choice probabilities for the models using the β-term that minimized the pair-wise Euclidean distance between the human and model's choice probabilities.

J. Benchmarking on OpenML-CC18

Table 7 contains the full set of results for all tasks and models.

K. Software and Data

We have made the data and code available under the following link: `https://github.com/akjagadish/ermi`

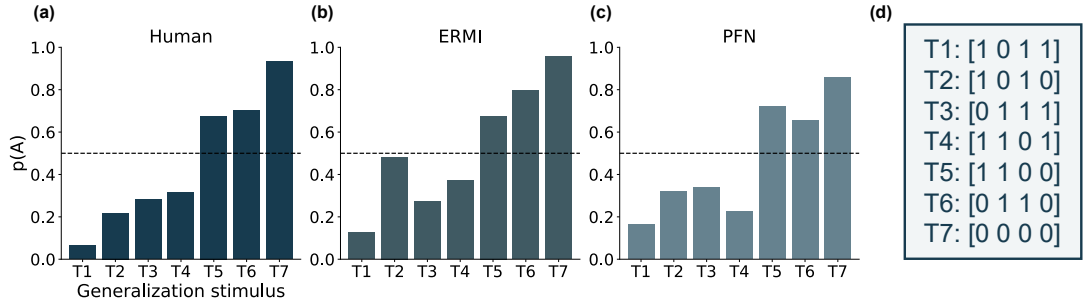

Figure 13. **Supplementary material accompanying Figure 4: (a-c)** Average categorization probabilities of transfer stimuli T1-T7 for (a) humans (b) ERMI (c) PFN. **(d)** The encoding scheme used for the seven transfer stimuli is provided for reference. Human data in (a) was reproduced from Johansen & Palmeri (2002). ERMI and MI were simulated on the same experiment for 77 runs, with inverse temperature settings that resulted in the lowest mean-squared error compared to humans, which was $\beta = 0.9$ for ERMI, and $\beta = 0.1$ for PFN.

Table 7. Detailed performance metrics of different models on OpenML-CC18 benchmarking suite.
Abbreviations: Software defect prediction (SDP) and Service-Center (SC)

DATA SET	LOG. REG.	SVM	XGBOOST	TABPFN	ERMI
KR-VS-KP	0.8257	0.8514	0.7986	**0.8664**	0.8450
CREDIT-G	**0.6421**	0.6357	0.6350	0.6036	0.6150
DIABETES	0.6771	**0.7079**	0.6786	0.6886	0.6950
SPAMBASE	0.5407	0.7664	0.7536	**0.7993**	0.7757
TIC-TAC-TOE	0.5536	0.5950	**0.6071**	0.5914	**0.6071**
ELECTRICITY	0.5543	0.6007	**0.7036**	0.6871	0.6436
PC4-SDP	0.7136	0.7521	**0.7886**	0.7707	0.7714
PC3-SDP	0.6514	0.7264	**0.7357**	0.7279	0.7107
KC2-SDP	0.5893	0.7314	**0.7257**	**0.7257**	**0.7257**
KC1-SDP	0.6271	0.6707	**0.6743**	0.6679	0.6521
PC1-SDP	0.5336	0.5964	0.6514	0.6064	0.6493
WDBC	0.9121	0.9207	0.9014	**0.9221**	0.9093
PHONEME	0.5793	**0.7314**	0.6979	0.6921	0.7200
QSAR-BIODEG	0.5779	0.7014	0.6850	0.6921	**0.7064**
ILPD	0.5493	**0.6386**	0.6229	0.6121	0.6286
OZONE-LEVEL-8HR	0.6614	0.6907	0.6707	0.6471	**0.6950**
BANKNOTE-AUTHENTICATION	0.7721	0.9229	0.8457	**0.9657**	0.9379
BLOOD-TRANSFUTION-SC	0.4714	0.5493	0.5879	0.5671	**0.6186**
PHISHING WEBSITES	0.7929	0.8071	**0.8157**	**0.8157**	0.8129
BANK-MARKETING	0.5829	0.5614	**0.7386**	0.7350	0.7171
WILT	0.5171	0.5736	0.6393	0.6371	**0.6507**
NUMERAI28.6	0.4857	0.4779	**0.5029**	0.4779	0.4986
CHURN	0.6321	0.7271	0.6800	0.7186	**0.7329**
MEAN ACC.	62.80 ± 0.66	69.29 ± 0.62	$70.17\% \pm 0.52$	$70.51\% \pm 0.63$	**70.95%± 0.54**
MEAN RANK	4.52 ± 0.21	2.76 ± 0.26	2.61 ± 0.30	2.85 ± 0.27	**2.26 ± 0.22**

2.5 Meta-learning ecological priors from large language models captures human learning and decision making

Jagadish, A. K., Thalmann, M., Coda-Forno, J., Binz, M.*, & Schulz, E.* (2025). Meta-learning ecological priors from large language models captures human learning and decision making.

Contributions in-context

Humans are remarkably adaptive, making good decisions in complex and uncertain environments. But where do these abilities come from, and how can we model them? This question has long challenged researchers. From Brunswik [134] to Gigerenzer [111], many influential psychologists have sought to answer this question. In recent times, two dominant computational frameworks have emerged to tackle this challenge: rational analysis [52] and ecological rationality [112]. Rational analysis derives optimal strategies within formal models of the environment, but its reliance on explicit assumptions on environmental properties means that it is constrained to relatively simple settings. Ecological rationality, in contrast, emphasizes heuristics tuned to real-world structure, offering greater flexibility but requires researchers to hand-craft adaptive strategies, making it difficult to scale to domains where effective heuristics are unknown (see also Section 1.5 and Section 1.7 in Background).

In this work, we introduce *ecologically rational analysis*, a new computational framework that unifies the normative foundations of rational analysis with the ecological grounding of heuristic strategies. By using LLMs to generate ecologically valid task distribution and training neural networks via meta-learning to solve these problems, we derived learning algorithms that are adapted to the statistics of real-world environments. The resulting models–*Ecologically Rational Meta-learned Inference (ERMI)*—are flexible, efficient, and capable of mirroring how people learn, categorize, and decide.

Specifically, we first showed that ERMI reproduces hallmarks of human function learning such as learning difficulty, learning speed, generalization, and biases toward specific functions (e.g. positive linear functions with zero offset). Its predictions also better captured human behavior than the rational model of function learning in two human function learning experiments [61]. Second, we demonstrated that ERMI captures qualitative and quantitative aspects of human category learning, from learning difficulty to learning strategy and generalization patterns, and outperforms eight established cognitive models in two different human experiments [135, 136]. Third, we showed that ERMI can flexibly shift between heuristics depending on the properties of the underlying task and accurately model human decisions in two additional experiments on heuristic decision making [22].

Finally, we outlined the current limitations of this approach, discussed potential extensions, and considered broader implications (see Section 3.1 and Section 3.2 in Outlook for extended discussion). Together, our results suggest that much of human learning and decision making may

be explained by adaptation to the statistical structure of the world around us.

Meta-learning ecological priors from large language models explains human learning and decision making

Akshay K. Jagadish[1, 2, 3, 4, *], **Mirko Thalmann**[2], **Julian Coda-Forno**[2], **Marcel Binz**[2,+], **and Eric Schulz**[2,+]

[1]Princeton University, Princeton, USA
[2]Institute for Human-Centered AI, Helmholtz Computational Health Center, Munich, Germany
[3]Eberhard Karls University of Tübingen, Tübingen, Germany
[4]Computational Principles of Intelligence, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[*]akshaykjagadish@gmail.com
[+]these authors contributed equally to this work

ABSTRACT

Human cognition is profoundly shaped by the environments in which it unfolds. Yet, it remains an open question whether learning and decision making can be explained as a principled adaptation to the statistical structure of real-world tasks. We introduce ecologically rational analysis, a computational framework that unifies the normative foundations of rational analysis with ecological grounding. Leveraging large language models to generate ecologically valid cognitive tasks at scale, and using meta-learning to derive rational models optimized for these environments, we develop a new class of learning algorithms: Ecologically Rational Meta-learned Inference (ERMI). ERMI internalizes the statistical regularities of naturalistic problem spaces and adapts flexibly to novel situations, without requiring hand-crafted heuristics or explicit parameter updates. We show that ERMI captures human behavior across 15 experiments spanning function learning, category learning, and decision making, outperforming several established cognitive models in trial-by-trial prediction. Our results suggest that much of human cognition may reflect adaptive alignment to the ecological structure of the problems we encounter in everyday life.

Significance Statement

Humans are remarkably adaptive, making good decisions in complex, uncertain environments. But where do these abilities come from and how can we model them? This work introduces a new approach that combines insights from psychology and machine learning to explain human cognition as an adaptation to ecological environments. By using large language models to generate realistic problems and training neural networks to solve them, we show that simple general-purpose systems can mirror how people learn, categorize, and decide. Our results suggest that much of human learning and decision making may be explained by attunement to the structure of the world around us.

Introduction

It is a truth universally acknowledged that a mind in search of a decision is influenced by its environment. Charles Darwin[1] showed that species are adapted to their environmental niche to survive. Egon Brunswik[2] proposed that people carefully interpret the signals in their surroundings to make judicious decisions. Herbert Simon[3] emphasized that human behavior is the result of the interplay between limited cognitive resources and the structure of the environment. Gerd Gigerenzer[4] furthered this notion by introducing the concept of ecological rationality, proposing that minds adapt to their environments by relying on simple context-specific strategies. Yet it remains unclear how attuned human learning and decision making are to the statistical structure of ecologically valid environments.

Two prominent frameworks have sought to address this question through computational modeling: rational analysis[5] and ecological rationality[6]. While rational analysis seeks optimal strategies within formal models of the environment, ecological rationality emphasizes heuristics tuned to the structure of real-world tasks. Although rational analysis offers a principled way to derive an adaptive strategy, it requires defining a formal model of the environment. This requirement limits its applicability to relatively simple environments. Ecological rationality, on the contrary, offers a flexible way to model real-world behavior, but it relies on the researcher to hand-design suitable heuristics. This reliance makes it challenging to extend the framework to new domains where effective heuristics have yet to be discerned.

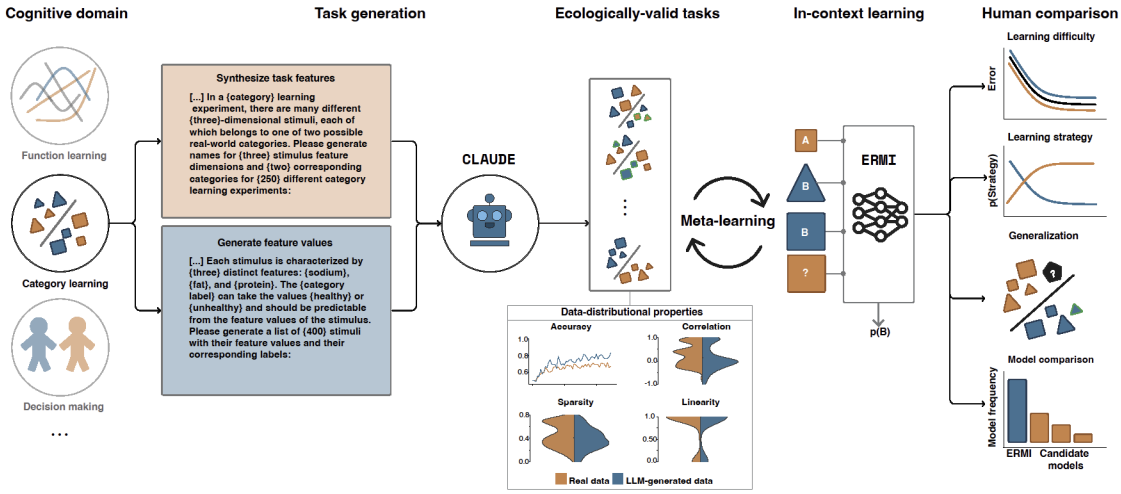

Figure 1. Schematic of Ecologically Rational Meta-learned Inference: Ecologically Rational Meta-learned Inference (ERMI) is domain-agnostic and can be applied to any cognitive domain. Let us consider category learning as the domain of interest for this illustration. The first step in deriving ERMI is to use a LLM (e.g., CLAUDE-V2) to generate ecologically valid tasks. Task generation from an LLM proceeds in two stages: first, the LLM synthesizes plausible task features (e.g., predicting whether a food item is healthy or unhealthy based on sodium, fat, and protein content); second, it generates corresponding input-target pairs consistent with these features[10]. Once a sufficient number of category learning tasks are generated, we analyze their distributional properties, such as classification accuracy, input feature correlation, sparsity in predictive features, and linearity of category structures, and compare them to real-world datasets (e.g., OpenML-CC18;[14]) to verify their ecological validity. We then derive computational models that internalize these ecological priors by training a neural network (e.g., transformer;[15]) on the LLM-generated tasks using meta-learning. This yields a family of in-context learners, termed Ecologically Rational Meta-learned Inference (ERMI), which flexibly adapt to the statistical structure of naturalistic problems. In category learning, the resulting models are evaluated against human behavior across four key dimensions: learning difficulty, learning strategy, generalization, and quantitative fit to human behavior through model comparison.

We introduce ecologically rational analysis, a framework that synthesizes the strengths of rational analysis and ecological rationality. This framework enables the automated derivation of computational models that implement approximately optimal strategies directly adapted to the statistical structure of natural environments. These models can subsequently be interrogated – through classical psychological experiments, for example – to elucidate how and which environmental properties give rise to human behavior.

To develop this framework, we draw on two recent advances in machine learning: large language models (LLMs) and in-context learning[7]. LLMs are generative models trained on internet-scale corpora, capable of capturing the statistical regularities that characterize real-world tasks and domains[8,9]. We harness this capacity to generate ecologically valid learning environments: problems that approximate the kinds of structure humans are likely to encounter in everyday life[10,11]. In-context learning refers to the ability of neural networks to learn from examples presented within a sequence, adapting their behavior purely through internal activations, without any parameter updates[7]. When derived via meta-learning, in-context learning has been shown to approximate Bayes-optimal inference conditioned on the statistics of its training distribution[12,13].

By meta-learning on tasks generated by LLMs, we develop models that internalize the ecological priors inherent in these environments. We term this class of models Ecologically Rational Meta-learned Inference (ERMI): a family of in-context learners that flexibly adapt to the statistical structure of naturalistic problems. We find that ERMI robustly captures human behavior across 15 experiments encompassing three core domains of cognition: function learning, category learning, and decision making. Beyond accounting for hallmark behavioral signatures within each domain, ERMI yields superior trial-by-trial predictions of human choices relative to a diverse array of established cognitive models. Collectively, these findings suggest that adaptive alignment with environmental statistics is sufficient to account for a broad spectrum of human learning and decision making behavior.

Results

In what follows, we describe ERMI and demonstrate how it can be used to model human learning and decision making across diverse cognitive domains; see Figure 1 for an overview.

The core idea behind ERMI is that adaptive behavior reflects the internalization of ecological priors. To capture these priors, ERMI uses LLMs as generative engines to construct ecologically valid tasks (see scalable generation of cognitive tasks from LLMs in Methods). This generation process involves two stages: first, the LLM proposes plausible task features (e.g., predicting weight from calories consumed); second, it generates corresponding input-target pairs for the given task features (e.g., specific calorie and weight values)[10]. Importantly, the generated targets correspond to ground truth values and *not* human predictions. By querying LLMs, we synthesize a rich and diverse set of cognitive tasks that approximate the distribution of problems found in natural settings.

ERMI then uses meta-learning[16–18] to derive computational models adapted to the LLM-generated tasks (see Methods). The resulting models implement approximately Bayes-optimal policies that adapt in-context – modifying internal activations rather than parameters – to the structure of encountered problems[12,13]. This approach allows us to systematically test whether optimal alignment with ecological statistics is sufficient to account for hallmark patterns of human learning and decision making.

In a series of experiments, we demonstrate how ERMI captures human behavior across three core domains of cognition: function learning, category learning, and decision making. For each domain, we show that it replicates core behavioral signatures and provides better trial-by-trial predictions of human choices compared to established cognitive models.

Function learning

Psychologists have been interested in understanding how people learn the functions underlying the association between an input and a target since the 1960s[19]. Much of these studies have focused on mapping a single-dimensional input to a response, called single-cue function learning[20,21], which is also the focus of this work.

In these tasks, participants observe input-output pairs, typically receiving feedback on the true response after each prediction. The underlying function is unknown and must be inferred through trial and error. Once trained, participants are tested on previously unobserved inputs within the range of their prior experience (interpolation) or outside that range (extrapolation). Previous work has revealed several hallmark findings: people interpolate more accurately than they extrapolate, sometimes performing as well on novel interpolated inputs as on the training set itself[22]; and they exhibit systematic biases when generalizing, favoring linear functions with positive slopes and minimal offsets[21,23,24].

ERMI provides a framework for probing the origins of these behavioral patterns. Following the procedure outlined previously, we first construct a dataset of about 10.000 one-dimensional function learning tasks designed to reflect the diversity of functional relationships found in natural environments (see Figure 2C for examples). We analyze the statistical properties of the LLM-generated tasks and compare the tasks with real-world regression problems[25]. We found that the generated functions comprised approximately 75% linear, 12% exponential, 7% quadratic, and 6% periodic relationships (see Figure 2A), a distribution that mirrors both environmental regularities and human difficulty rankings in function learning[26]. A model-based analysis of linear fits revealed a pronounced bias toward positive slopes with near-zero offsets (see Figure 2B), consistent with known human biases in extrapolation[27,28].

Then, we asked to what extent ERMI can replicate the characteristic patterns of human function learning? Drawing on prior work[26], we focused on five well-established findings in function learning: (i) linearly increasing functions are learned more readily than decreasing functions[30–32]; (ii) linearly increasing functions are also learned faster than nonlinearly increasing functions[22,33]; (iii) monotonic functions are learned more effectively than non-monotonic ones[19,30,33]; (iv) cyclic functions are more challenging than non-cyclic functions[33]; and (v) generalization is more accurate in interpolation than in extrapolation[19,22,34].

To test whether ERMI reproduces these patterns, we sampled functions from each class and assessed the model's learning dynamics. We measured learning speed and accuracy using the rate of change and mean-squared error (MSE) across trials (see Methods for details). Strikingly, ERMI mirrored human behavior across all five phenomena: (i) it learned positive linear functions (M_{MSE}=0.5260, SEM=0.0043, t=-70.1587, $p<0.001$) better than negative linear functions (M_{MSE}=0.9339, SEM=0.0039); (ii) it grasps linear functions with positive slopes more rapidly, reaching minimum MSE in fewer trials (M_{trial}=13.29, SEM=0.051) compared those with negative slopes (M_{trial}=14.22, SEM=0.048; t=-13.38, $p<0.01$); (iii) it mastered monotonic functions (M_{MSE}=0.8895, SEM=0.0028, t=-145.5417, $p<0.01$) more accurately than non-monotonic ones (M_{MSE}=1.5255, SEM=0.0032); (iv) it learned non-cyclic functions (M_{MSE}=1.0423, SEM=0.0025, t=-89.6918, $p<0.01$) more readily than cyclic ones (M_{MSE}=1.5508, SEM=0.0054, t=-89.6918, $p<0.01$); and (v) it achieved better prediction performance during interpolation (MSE=0.0017;[27]) than during extrapolation (MSE=0.0022); see Figure 2D-E.

A well-established finding in the function learning literature is that humans tend to underestimate functional relationships during extrapolation, particularly for linear functions, with a characteristic bias toward zero offset[27]. To assess whether ERMI exhibits a similar pattern, we conditioned the model on input-target pairs sampled from a linear function, following the

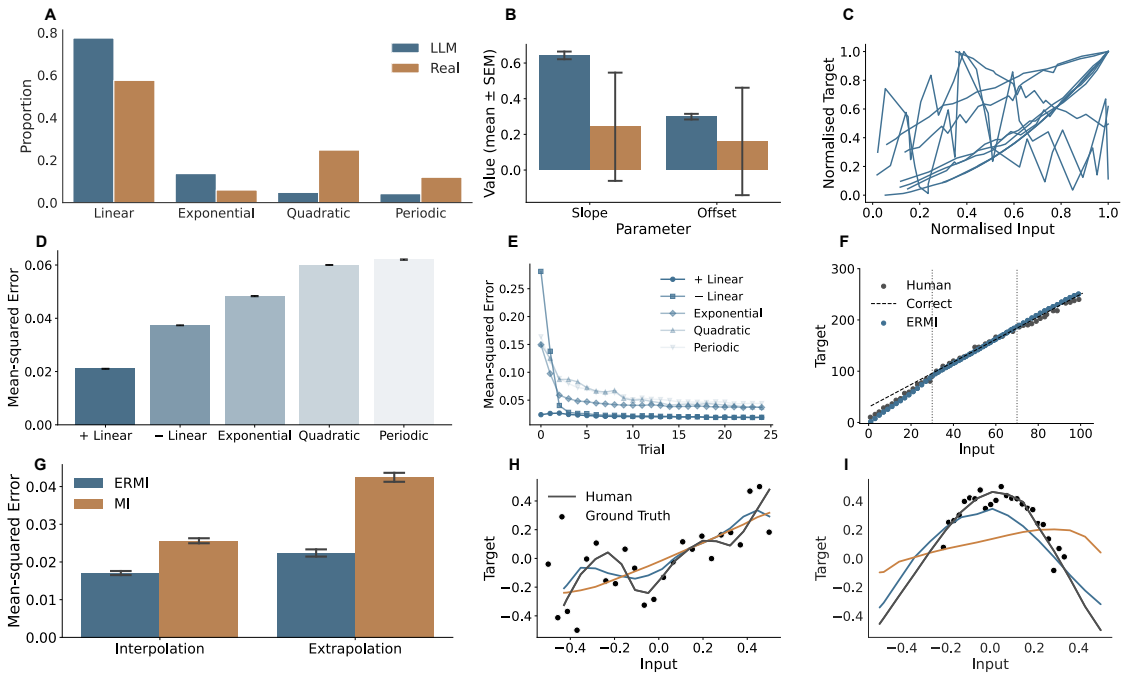

Figure 2. Function learning: **A** Proportions of different function types in real-world datasets[25] and LLM-generated datasets. **B** Parameters for slope and offset for linear functions fitted to both datasets. **C** Example functions sampled from the LLM-generated datasets. **D** Mean-squared error (MSE) of ERMI when simulated on five function types, namely, positively-sloped linear, negatively-sloped linear, exponential, quadratic, and periodic functions (mean over all runs and trials). **E** MSE of ERMI for the five function types mentioned above unrolled over trials (mean over all runs). **F** Simulations of ERMI on linear function from[27] along with human predictions extracted from the original plot (gray lines mark interpolation range between 30 to 70 and extrapolation region between 0 to 30 and between 70 to 100). **G** MSE during interpolation (Experiment 2 of original study) and extrapolation (Experiment 4) for ERMI and Meta-learned inference (MI) with hand crafted prior when simulated on tasks from[29]. **H** Representative example for interpolation. **I** Representative example for extrapolation.

procedure of Kwantes and Neal[27] (see human studies in SI for additional details), and simulated its predictions for input values outside the training range. We compared ERMI's responses to human data in both interpolation and extrapolation regimes. For human responses, we drew on data from Experiment 2 of Kwantes and Neal[27], in which participants directly estimated numerical values for a given input – an evaluation method that closely parallels our procedure for ERMI. We found that, like humans, ERMI systematically underestimated responses in the extrapolation range, with a stronger bias in the lower region (MSE=0.0043; 0-30 on the x-axis, which is marked using grey lines in the Figure 2F) than in the upper region (MSE=0.0003; 70-100 on the x-axis). It showed a bias towards zero offset comparable to that observed in human participants (see Figure 2F). In addition, ERMI's predictions agreed better (MSE=0.0002) with human responses than a meta-learned (MI; MSE=0.00054) with hand-crafted prior used by Lucas and colleagues[35]; see Methods for details.

Beyond qualitative signatures, a critical test of any model is whether it can capture the fine-grained structure of human behavior at the trial level. To this end, we evaluated ERMI on the function estimation task introduced by Little and colleagues[29]. In this task, participants viewed 24 scatter plots, each depicting data from a different fictional scientific experiment, and were asked to draw what they believed to be the true underlying causal function. The scatter plots were generated from linear, quadratic, or cubic polynomial functions with added Gaussian noise. Crucially, the scatter plots were presented in two distinct formats. In the zoomed-in condition, data points filled the entire plotting area, enabling assessment of interpolation. In the zoomed-out condition, data points were centrally located and occupied only 40% of the plot area, encouraging participants to extrapolate beyond the range of the observed data; see human studies in the SI for additional details.

We conditioned ERMI on the same training data shown to participants and generated predictions for the identical input values. To quantify model fit, we computed the MSE between ERMI's predictions and human responses across all test inputs. As shown in Figure 2G, ERMI provided a closer fit to human judgments than MI in both interpolation and extrapolation conditions. Specifically, ERMI achieved a lower MSE (M_{MSE}=0.0171, SEM=0.0014, t = -9.9944, $p < 0.01$) compared to MI (M_{MSE}=0.0256, SEM=0.0020) during interpolation, and likewise outperformed MI during extrapolation (ERMI: M_{MSE}=0.0223, SEM=0.0018; MI: M_{MSE}=0.0424, SEM=0.0034, t = -13.1479, $p < 0.01$). For illustrative examples, Figure 2H-I shows ERMI's predictions, alongside those of MI, for interpolation and extrapolation condition from a representative participant; see Figure S2 in the SI for additional examples. Together, these findings suggest that a rational model attuned to the statistics of ecologically valid function learning problems is sufficient to capture much of human function learning.

Category learning

To examine whether these findings generalize, we turned to a second domain that has been widely studied in cognitive psychology, namely category learning[36]. In a typical category learning task, participants are presented with inputs sequentially and must assign each one to a set of known categories. After each trial, they receive feedback that indicates whether their classification was correct. The underlying rule that governs category membership – referred to as the category structure – is hidden from participants and must be inferred through trial and error. During the test phase, participants classify both previously seen (training) and novel (transfer) input without receiving feedback. This design allows for simultaneous assessment of learning performance on familiar examples and generalization to new instances.

It has been observed that humans find learning certain category structures more difficult than others[37]. Furthermore, the categorization strategy they use changes during the course of the experiment from an exemplar-based strategy to a prototype-based strategy[38]. In addition, the way they generalize to unseen inputs is systematic, following a rule-plus-exception-based model[39].

To investigate the role of ecological adaptation in explaining these findings, we again turn to ecologically rational analysis. Following a previously established procedure, we generated about 10.000 category learning problems and inspected their underlying statistics. Like in function learning, we found that LLM-generated category learning problems capture key statistical properties of real-world classification datasets[14]. Specifically, we observe that (i) the generated category learning problems are noisy, yet classifiable, like real-world classification problems (Figure 3A); (ii) they contain inputs whose features exhibit a full range of correlations, from non-existent or low values to nearly complete correlation, as in real-world tasks (Figure 3B); (iii) only a few feature dimensions within each task substantially contribute to classification, indicating sparsity in the feature space commonplace in real-world tasks (Figure 3B; larger Gini coefficient values indicate higher sparsity); and (iv) the category structure observed in the generated classification problems is predominantly linear akin to real-world tasks (Figure 3B; higher values indicate more linearity). These findings confirm the ecological validity of LLM-generated category learning tasks.

Moving to the next step, we derived ERMI by meta-learning on these category learning problems and evaluated how well it can capture various aspects of human category learning. To explain human learning difficulties during category learning using ERMI, we consider the study by Shepard et al.[37]. In this study, the authors considered six different category structures (labeled TYPE 1 to TYPE 6). In TYPE 1 problems, all items in a category share one particular feature value (e.g., they are all black), TYPE 2 problems are defined by a combination of two feature values (i.e., XOR problems), TYPE 3-5 problems combine a rule with exceptions, and TYPE 6 problems require the memorization of individual items as rules and the similarity to other

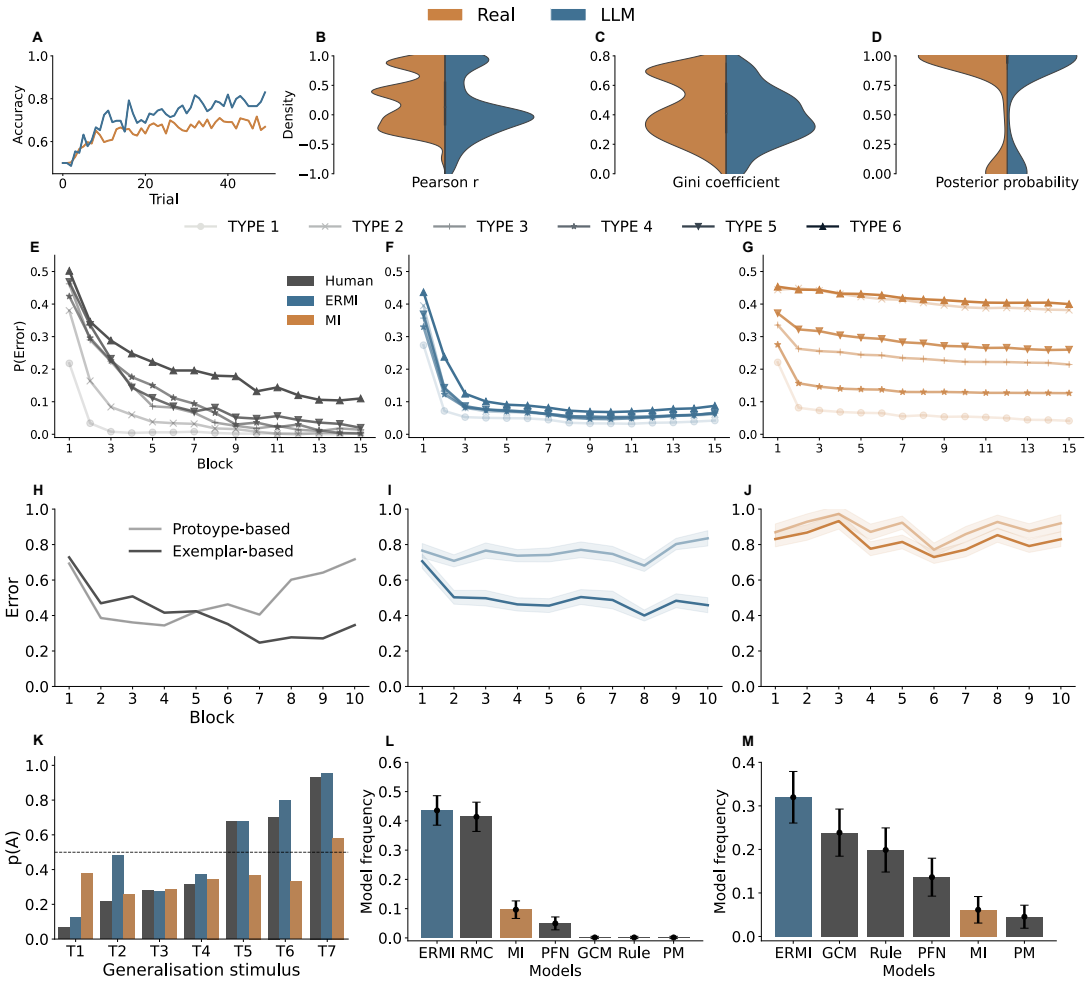

Figure 3. Category learning:**A** Mean task performance of a logistic regression model over trials for real-world classification tasks[14] in orange and LLM-generated tasks in blue. **B** Density plot of Pearson's correlation coefficients between feature dimensions. **C** Gini coefficients over logistic regression weights, which provides a measure of sparsity (high values indicates greater sparsity). **D** Posterior probability measuring the linearity of category learning tasks. **E–G** Average error probabilities for each task TYPE across 16-trial blocks for **E** humans (in gray), **F** ERMI (in blue), and **G** MI (in orange). Human data in **E** reproduced from Table 1 in[40]. ERMI and MI were simulated on TYPE 1–6 tasks for 50 runs using the inverse temperature (β) that minimized mean-squared error with respect to human data: $\beta = 0.4$ for ERMI and $\beta = 0.9$ for MI. **H–J** Average error of exemplar- and prototype-based models fitted to **H** human choices, **I** simulated choices from ERMI, and **J** simulated choices from MI across 56-trial blocks. Human data in **H** reproduced from[38]. ERMI and MI were simulated using inverse temperature values fitted to participants' choices in[41]; ERMI (M_β=0.09, SEM=0.01) and MI (M_β=0.17, SEM=0.02). Shaded regions indicate standard error of the mean. **K** Average categorization probabilities of transfer inputs T1–T7 for humans (gray), ERMI (blue), and MI (orange). Human data reproduced from[39]. ERMI and MI were simulated on the same experiment over 77 runs using the best-fitting inverse temperatures: $\beta = 0.9$ for ERMI and $\beta = 0.1$ for MI. **L** Posterior model frequency of participants' choices in[42] across seven computational models. **M** Posterior model frequency of participants' choices in[41] across six computational models.

items is not informative; see human studies in the SI for details. The task difficulty of the six problem types increases from 1 ($M_{p(Error)}$ = 0.0201) to 6 ($M_{p(Error)}$ = 0.2048), as shown in Figure 3E. The error rates for TYPE 2-5 problems fall between those of TYPE 1 and TYPE 6; see Table S3 in SI for details. ERMI – when simulated on tasks from the Shepard et al. study[37] – displayed learning curves that are difficulty dependent and follow the same ordering as people's; see Figure 3F. Quantitatively, ERMI (MSE = 0.03) captured human learning difficulties better than meta-learned inference with a hand-crafted prior (MI; MSE = 0.26); see Methods for details.

We then investigated whether ERMI also captures human trial-by-trial choices during category learning by considering a replication of the original Shepard's study[37] by Badham et al.[42]; see human studies in the SI for details. We performed a Bayesian model comparison between ERMI and five other computational models, which included the rational model of categorization (RMC;[43]), a prototype model (PM;[44]), an exemplar model (generalized context model (GCM);[45]), a rule-based model (Rule;[34] and a meta-learned inference model with Bayesian logistic regression prior (MI) and Bayesian neural network prior (PFN;[46]); see Methods for details on the baseline models, as well as the model fitting and comparison procedure. We found that in terms of the posterior model frequency (PMF), which measures how often a model offers the best explanation in the population, ERMI explains human choices more frequently (M_{PMF}=0.43, SEM=0.05) compared to the other models, with the RMC coming in close second (M_{PMF}=0.41, SEM=0.05); see Figure 3L.

After that, we tested whether ERMI shows a similar shift in its categorization strategy as humans[38]. In this study, participants classified 14 six-dimensional inputs into two categories. These categories were assigned based on a nonlinear decision rule; see human studies in the SI for details. The authors then fitted a prototype and an exemplar model to the observed behavior and found that the prototype model better explained the people in the early blocks but in the later blocks, their choices aligned more closely with the exemplar model, as shown in Figure 3H. When simulated on tasks from the same study, we found ERMI's strategy to be indistinguishable between prototype-based and exemplar-based in the beginning of the experiment, but with experience, it became increasingly more exemplar-based as observed in humans (see Figure 3I). In contrast, MI does not display such a transition in strategy, as shown in Figure 3J. Furthermore, we compared ERMI with other competing models in the prediction of human choices at the trial level, for which we used human data from a replication of the original study by Devraj et al.[41]. As shown in Figure 3M ERMI (M_{PMF}=0.32, SEM=0.06) predicted human choices the most frequently, followed by the GCM (M_{PMF}=0.24, SEM=0.05) and the rule-based model (M_{PMF}=0.20, SEM=0.05).

Finally, we examined whether ERMI displays the same generalization patterns as people when they observe inputs not part of the training phase[39]. In the training phase of this study, participants performed binary classification of nine four-dimensional inputs. Subsequently, in the test phase, they were probed on seven transfer inputs (labeled T1-T7; see Methods for their encodings) for which they did not receive any feedback; see human studies in the SI for details. The latter was intended to examine how they would generalize the learned category structure to unseen inputs; see Method for details. Figure 3K shows the proportion of responses in which the participants assigned category A to the seven transfer inputs (in gray). It can be seen that participants assigned the transfer inputs T5, T6, and T7 mainly to category A and the inputs T1, T2, T3, and T4 mainly to category B. ERMI – when evaluated on the same task – generalizes to unseen inputs in a human-like way by classifying the inputs T1, T3, and T4 more often as category B and the inputs T5, T6, and T7 more often as category A; see Figure 3I (in blue). The only deviation from human-like generalization is input T2. Although ERMI classified it as category A at the chance level, the participants predominantly assigned it to B. We speculate that this is because T2 resembles the category prototype along two only dimensions, while other inputs categorized as B matched along three dimensions. Yet again, MI did not show the same pattern as in humans, both qualitatively (Figure 3K; in orange) and quantitatively, with the Euclidean distance between the choice probabilities of humans and ERMI (0.29) being lower than between humans and MI (0.67).

These results indicate that in addition to capturing human function learning, ERMI also captures human category learning.

Decision making

The question of how people decide between multiple options and how they improve on it with experience has been extensively studied in economics[47], psychology[4], and neuroscience[48]. Does ERMI also extend to this domain?

For our analyses, we considered the paired comparison task[49]. Participants in this task decide between two options, each of which is characterized by different feature values. The feature values are associated with a value on an unobserved criterion and participants have to learn which option has the higher criterion. We consider a sequential variant where participants take decisions one at a time with feedback provided on which option had a higher criterion value after each trial.

The strategies people use to make decisions in a paired comparison task are widely contested. While economists have taken a rational perspective that suggests that people weigh the different cues appropriately while making decisions[47,50], proponents of ecological rationality have argued that people are incapable of such reasoning due to their cognitive constraints[51,52]. Instead, they have proposed the view that people rely on simple heuristics, which are short-cut strategies that produce competitive performance despite using only parts of the available information[53]. Recent work[49] has shown that people adopt different decision-making heuristics depending on the structure of paired comparison tasks[54]. When participants knew the importance of

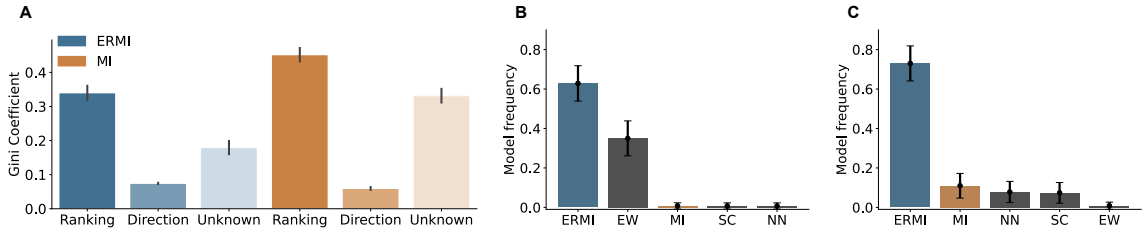

Figure 4. Decision making: **A** Mean Gini coefficients computed over weights produced ERMI and MI on the tenth trial of the paired comparison tasks sampled from the generative model used in the Binz and colleagues[49] for three conditions: ranking, direction and unknown. **B** Posterior model frequency of participants' choices from experiment 3b of Binz et al. study[49], which uses four attributes for each option. **C** Posterior model frequency of participants' choices from experiment 3a of Binz et al. study[49], which uses two attributes for each option.

attributes to the criterion but not the direction of their correlation with it (ranking condition), they used a one-reason decision strategy. When the direction was known but not the ranking (direction condition), they relied on an equal weighting strategy. Finally, when neither ranking nor direction was known (unknown condition), they used a weighted combination of attributes to guide their choices.

To further examine how data distributional properties influence heuristic choice, we adapted the previously described procedure to generate three sets of problems, each containing approximately 7.000 tasks, reflecting one of three conditions: ranking, direction, and unknown. This was done using three condition-specific prompts specifying: (i) that attributes be rank-ordered by importance to the target; (ii) that attributes correlate positively with the target; or (iii) no additional information to allow free-form generation. We then verified that the generated tasks matched their intended conditions: tasks in the ranking condition showed more rank-ordered feature weights than those in the unknown condition, and tasks in the direction condition had features more strongly (positively) correlated with the target (see Figure S7 in the SI). After that, we constructed paired comparison trials by randomly sampling two options from each dataset and pitting them against each other. We then derived an ERMI model by meta-learning on LLM-generated tasks for each condition.

The resulting ERMI models were first simulated on decision making problems from Binz and colleagues[49]. To examine the strategy being implemented by ERMI, we computed the Gini coefficient over attribute weights produced by ERMI (see Methods for details). Higher values for the Gini coefficient indicate more sparse weights, which corresponds to a one-reason decision making strategy, lower values correspond to equally weighted attributes, and values in between correspond to a weighted-additive strategy. As shown in Figure 4A, we found that ERMI uses the same heuristics that people use in the respective condition. That is, ERMI trained on decision making problems from the ranking condition implements a one-reason decision making strategy (M_{Gini} = 0.3399, SEM = 0.0631), in the direction condition it uses an equal weighting strategy (M_{Gini} = 0.0743, SEM = 0.0138), and in the unknown condition it relies on a weighted combination of attributes (M_{Gini} = 0.1794, SEM = 0.0333). These results are also consistent with the strategies used by meta-learned inference (MI) with a hand-crafted prior for each condition; see Methods for details.

In addition to the simulation study, we evaluated whether ERMI explains human decision making better than competing models by conducting a model comparison on human data from Binz et al.[49]. We considered human data from two experiments, one with options containing two attributes and the other with four, in which participants performed decision-making tasks without receiving any side information (see human studies in SI for details). Compared to other baseline models – namely, single-cue decision making (SC), equal-weighted strategy (EW), feedforward neural network (NN), and MI – ERMI accounts for human responses most frequently in both the two-attribute experiment (M_{PMF}=0.7299, SEM=0.0888) and the four-attribute experiment (M_{PMF}=0.6284, SEM=0.0897). These results suggest that ERMI converges to the same decision making strategies that people have been shown to use, with the particular strategy shaped by the statistical structure of decision-making problems.

Discussion

In the late 1990s, Gerd Gigerenzer and his colleagues conducted what, in hindsight, stands as one of psychology's most endearingly simple yet profoundly revealing studies. They asked participants whether they recognized the names of various cities or companies—and found that when people chose the company they recognized between two options, their choices reliably predicted which company had higher stock returns. This surprisingly effective strategy, dubbed the recognition heuristic, is a lexicographic decision rule similar to the ones we modeled that stops at the first discriminating cue—in this case, recognition. But how could such a seemingly naive rule succeed in complex contexts like financial forecasting? Gigerenzer proposed that

the frequency with which people encounter names in everyday life –on television, in conversations, or in headlines– contains statistical signals informative for decision making[55], a hypothesis he substantiated by meticulously counting company name occurrences in major newspapers.

Yet this insight raised a far-reaching question: could one ever scale ecological rationality beyond a single heuristic to explain the full complexity of human learning and decision making? After all, manually tallying newspaper mentions might work for isolated cues, but becomes hopelessly unwieldy when faced with the rich environments people face daily. How could one map the statistical fingerprints of vast environments across countless domains of cognition?

In recent years, an unexpected answer may have emerged: large language models. Indeed, it has been argued, by Alison Gopnik and colleagues[56], that LLMs, trained on the sprawling archive of human culture, can be seen as "cultural technologies," artifacts that distill the collective knowledge of societies. Where earlier researchers scraped headlines to estimate how often a name appeared in people's environments, LLMs now embed billions of such frequencies and co-occurrences, capturing statistical regularities on a scale previously unimaginable. This extraordinary capacity opens, for the first time, an opportunity to massively scale up ecological rationality. We can use LLMs to generate ecologically grounded tasks that reflect the natural statistics of human environments and test whether human learning and decision-making align with ideal inference under such ecological priors.

In the current work, we introduced ecologically rational analysis, a framework that leverages meta-learning and LLMs to do exactly that, i.e. to extend the logic of ecological rationality beyond individual heuristics and into the broader structure of human cognition. In particular, we developed a new class of models—called ERMI—that allowed us to investigate whether human learning and decision making approximate "ideal statistical inference under the structure of natural tasks and environments"[57]. Across 15 experiments spanning three core domains of human cognition, we found that ERMI can account for a substantial amount of variance in human behavior. Not only did ERMI capture key behavioral signatures in each domain, but it also provided superior trial-level prediction of human choices relative to established cognitive models. Taken together, these findings demonstrate that rational adaptation to ecologically valid task statistics is sufficient to account for much of human cognition.

A key strength of ERMI lies in its ability to derive priors and distill them into computational models without extensive hand-engineering. In contrast, traditional rational analysis requires researchers to manually specify the underlying data-generating distribution. For example, in the rational model of function learning, Lucas et al.[35] assumed a linearity-based prior but acknowledged uncertainty about its alignment with naturalistic environments, noting that "it is not realistic to directly measure the statistical structure of the environment, that is, what functions are truly more or less common"[35]. ERMI circumvents this issue by using LLMs to directly generate tasks with ecological statistics. Alternatively, ecological rationality often relies on researchers to manually construct heuristics that are "applicable to specific decision tasks and in particular domains—different tools for different tasks"[58]. By leveraging meta-learning to automatically derive computational models adapted to these ecological statistics, ERMI eliminates the need for hand-designing task-specific heuristics prevalent in ecological rationality framework.

Yet one may ask: Why not directly use an LLM to model human behavior, instead of meta-learning on LLM-generated tasks? We evaluated LLMs as direct behavioral models and found that ERMI consistently outperformed them in explaining human data (see Figure S5 in the SI), highlighting that LLMs do not capture human behavior out-of-the-box. Furthermore, even a strong alignment of LLMs with human behavior would not clarify why they are good models, given that they are trained on vast and opaque datasets that span human conversations, code, and various cultural artifacts that are difficult to analyze. ERMI, on the other hand, leverages LLMs solely as generative sources for ecological tasks used in meta-learning, allowing us to test a specific hypothesis about human cognition. Of course, what features of the environment are required to explain behavior in a particular domain still needs to be determined statistically, which re-introduces a degree of manual analysis that ERMI does not yet fully automate. In that sense, while ERMI reduces the burden of handcrafting heuristics and priors, it does not eliminate the need for scientific judgment in feature selection and interpretation. However, this remaining bottleneck also presents a new opportunity: by systematically varying environmental features across LLM-generated tasks and analyzing their impact on model fit and human behavior, we can begin to reverse-engineer the ecological ingredients that most shape cognition.

Future work should determine which additional components are required to account for human behavior beyond ecological rationality. ERMI offers a flexible foundation for integrating such components. First, we can incorporate participant-specific information into the data generation process, followed by meta-learning on these tailored datasets. This approach would enable personalized ERMI models that capture individual differences, particularly those shaped by environmental and demographic factors unique to each participant. Second, while computational models derived via ERMI currently emphasize adaptation to the environment, they largely ignore the role of cognitive constraints. Incorporating limits on computational complexity –such as attention, working memory, or representational capacity– could help explain additional variance in human behavior, especially in cases where people systematically deviate from ideal inference[49,59,60]. Notably, such constraints can either be explicitly modeled within the meta-learning process or may already be implicitly embedded in the training data itself, given that these data are generated by humans who are inherently resource-bounded. Third, ERMI could serve as a starting point for

fine-tuning on human choice data, following recent approaches[61–63]. This would allow for principled estimation of residual variance in behavior not yet captured by ERMI and help identify the cognitive mechanisms needed to close that gap.

Humans excel at learning and decision making in complex and uncertain environments. Our findings suggest that these cognitive abilities emerge largely through attunement to ecological structures. By harnessing ecologically rational analysis, combining psychology and machine learning, we demonstrate how general-purpose models, trained in realistic ecological tasks, can mimic much of human behavior. Looking ahead, we speculate that scaling up this approach to open-ended, embodied environments[64], which require processing high-dimensional visual information and executing complex control sequences, holds promise for expanding ecologically rational analysis. By leveraging multimodal foundation models to generate personalized, ecologically valid tasks and meta-learning for distilling those priors into adaptive computational models, we can systematically quantify how much of human behavior can be explained as an adaptation to previously encountered task structures and environments. If successful, this would significantly broaden the explanatory power of cognitive models, offering nuanced insights into the ecological roots of human cognition.

Methods

In this section, we first describe how we generate cognitive tasks at a scale that is sufficient to train in-context learning models from scratch and how we verify their ecological validity. Following this, we discuss how to derive in-context learning models via meta-learning and present other domain-specific cognitive models used as baselines. We then describe (a) how we simulate behavior from different models on these experiments, and (b) how we fit and compare these models to human data.

Scalable generation of cognitive tasks from LLMs

Generating cognitive tasks from an LLM entailed a two-stage process. In the first stage, we query an LLM to synthesize the names for input features and targets. For instance, an example input feature in function learning could be CALORIE INTAKE whose corresponding target is WEIGHT. In the second stage, the LLM is queried again but this time to generate numerical values for a given input feature and target pair generated from the first stage. That is, the LLM is tasked to generate different values for [CALORIE INTAKE, WEIGHT], for instance, [2300, 152.0] or [1850,143.0].

Below, we provide the prompts used in the two stages for the function learning domain; see SI for prompts for other domains. We used the following prompt to synthesize names for input features and targets for function learning:

Synthesize input feature name and its target

I am a psychologist who wants to run a function learning experiment. In a function learning experiment, a real-world feature is mapped to its corresponding target, with both feature and target taking on continuous values.

Please generate names for features and its corresponding target for **250** different function learning experiments:

– feature name, target name

Next, we prompted the LLM to generate values for a function learning task generated from the first stage:

Generate values for a given function learning task

I am a psychologist who wants to run a function learning experiment. For a function learning experiment, I need a list of features with their corresponding target. The feature in this case is **calorie intake**. The features take on only numerical values and must be continuous. The target, **weight**, should be predictable from the feature values and must also take on continuous values.

Please generate a list of **20** feature-target pairs sequentially using the following template for each row:

– feature value, target value

We generated a dataset containing around 10000 different function learning tasks with each task consisting of 20 data points from CLAUDE-V2[65]. The temperature parameter was set to one to induce diversity, and all other parameters were set to their default values. We chose CLAUDE as it can process up to 100.000 tokens, is instruction-tuned, cost-effective, and performed well out of the box on most of our preliminary analyses; see SI for information about the other LLMs we considered.

To use and analyze the generated cognitive tasks, we parse all necessary quantities from the output text from the LLMs using regular expressions and stored them in numerical format in comma-separated-value (csv) files. These stored csv files

are the datasets we use for further analysis. We expand on the parsing expressions, data-processing steps, and also provide a qualitative analysis of synthesized input feature and target names (see Figures S1, S3, S4, and S6) in the SI.

Verifying the ecological validity of LLM-generated cognitive tasks

To test the ecological validity of the generated cognitive tasks, we resort to two approaches. We either compare certain key statistics between LLM-generated tasks and a real world baseline, whenever we have access to a reasonable dataset, or compare it to real world statistics expected or predicted by prior work. We will discuss these tests for each domain individually below.

Function learning.

We compared the data distributional properties of the LLM-generated function learning problems with 60 real-world regression tasks curated by Lichtenberg and colleagues[25]. We downsampled all tasks in the dataset to a single input dimension by applying univariate feature selection using the F-statistic for regression, as implemented in SCIKIT-LEARN[66], and included only tasks without missing values and with at least one valid input dimension in our analysis. Note that each dataset was split into separate tasks of 25 datapoints each, yielding a collection of regression problems with fixed size for analysis.

For both real and LLM-generated function learning tasks, we estimated the relative frequency of different function classes within the dataset. We did this by first fitting models of different function classes to each LLM-generated task and then, assigning the function class with the best fitting model to the given task.

Specifically, we considered models from four well-studied function families, namely, linear, exponential, quadratic, and sinusoidal, from the literature[26]; see SI for their exact model instantiations. The parameters of these models, ϕ, were fit to data from the task to minimize the sum of squared errors (SSE) using the curve fit function from the SCIPY optimization library[67]. We then computed the Bayesian Information Criterion (BIC) for the fitted models from each function class, compared them against each other, and assigned the label of the function class that won the model comparison to a given task. Assuming $\hat{y}(\phi)$ and y correspond to predicted target from the fitted model with parameters ϕ and true target, respectively, BIC computation entailed the following steps:

$$\begin{aligned} \text{SSE} &= \min_{\phi} \sum_{i=1}^{N} (y_i - \hat{y}_i(\phi))^2 \\ \text{BIC} &= N \cdot \ln(\text{SSE}) + |\phi| \cdot \ln(N) \end{aligned} \tag{1}$$

where $|\phi|$ is number of parameters in a given model parameters, and N is number of data points per task. This SSE-based approximation of the BIC assumes that model errors are Gaussian with constant variance, under which the negative log-likelihood is proportional to the sum of squared errors.

We obtained the proportion of different function classes by computing a histogram over the assigned class labels for all tasks in a given dataset. Furthermore, we assessed whether the fitted slope term of the linear model were predominantly positive and whether the fitted offset term, from the same model, was close to zero.

Category learning. We compared the data distributional properties of the LLM-generated category learning tasks with a real-world classification benchmark[68]. For this, we used the OpenML-CC18 benchmarking suite, a curated collection of real-world classification tasks[14]. We downsampled all tasks in the OpenML-CC18 benchmark to four feature dimensions by applying univariate feature selection using the ANOVA F-test implemented in SCIKIT-LEARN[66] and included only binary classification tasks without any missing features in our analysis – amounting to 28 tasks.

We analyzed these collections of tasks in terms of their learning curves, input feature correlations, sparsity of predictive features, and linearity of the category structure. We obtained the learning curves by fitting a logistic regression model on a trial-by-trial fashion. For input correlations, we computed Pearson's correlation coefficient between every pair of features in the task. To get an estimate for task sparsity, we fitted a logistic regression model on the full data for each task and analyzed the sparsity of the resulting regression weights $\mathbf{w} \in \mathbb{R}^d$ using the Gini coefficient G:

$$G(\mathbf{w}) = \frac{\sum_{i=1}^{d} \sum_{j=1}^{d} |\mathbf{w}_i - \mathbf{w}_j|}{2d \sum_{i=1}^{d} \mathbf{w}_i} \tag{2}$$

For determining the linearity of the category structure, we fitted a logistic regression model and a logistic regression with second-order polynomial features on the full data $\mathscr{D}$ from each task. We then computed the BIC for both models and used them

to approximate the posterior probability that the linear model offers a better explanation of the data (assuming a uniform prior over models), see Equation 3.

$$p(M = \text{linear}|\mathscr{D}) \approx \frac{\exp(-0.5 \cdot \text{BIC}_{\text{linear}})}{\sum_{m \in \{\text{linear, polynomial}\}} \exp(-0.5 \cdot \text{BIC}_m)} \tag{3}$$

Decision making. We examined the distribution of input feature correlation, sparsity in predictive features, rank ordering of feature importance, and directionality of the features for the three LLM-generated decision making datasets belonging to ranking, direction, and unknown condition; see SI for details about their generation. For baseline, we considered the LLM-generated dataset in the unknown condition, as it allows contrasting dataset from the rank and direction condition with one that lacks explicit manipulation. See SI for data-distributional properties of LLM-generated decision making tasks.

For measuring correlation between input features, we compute pair-wise Pearson's correlation coefficient, following the same procedure we used in the domain of category learning; see Figure S7 (first column) in the SI for visualization.

To measure sparsity of task features, we followed the same procedure as in the category learning task but instead of a logistic regression model, we fitted a linear regression model that predicts the continuous valued targets from the task features; see Figure S7 (second column) in the SI for visualization.

For examining the rank ordering of feature importance, we fit a linear regression model, predicting the target from the input features. We then identified the feature with the highest absolute regression coefficient for each task and performed histogram over these positions to assess how often each feature was most predictive. If the intended manipulation was successful, we expect that the first feature should most frequently have the largest coefficient, followed by the second, and so on, reflecting a consistent ordering of feature relevance; see Figure S7 (third column) in the SI for visualization.

We assessed the directionality of each feature by examining the sign of the fitted regression coefficients from linear models as described above. If all coefficients are positive, it suggests that the intended manipulation was successful; see Figure S7 (fourth column) in the SI for visualization.

Ecologically Rational Meta-learned Inference

Having generated and tested the ecological validity of LLM-generated cognitive tasks, we then trained transformer-based models on those tasks to derive explicit in-context learning models adapted to the ecological task distribution. For this, we let a transformer-based model[15] auto-regressively predict a target, y_t which can either be a discrete category or a continuous response, for a given input, x_t, conditioned on all preceding input-target pairs, $(x_{1:t-1}, y_{1:t-1})$. After the model predicts targets for all inputs in the sequence, the parameters of model, θ, is updated based on the following objective:

$$\ell = \sum_t -\log p_\theta \left(y_t \mid x_{1:t}, y_{1:t-1}\right) \tag{4}$$

where p_θ defines the output probabilities produced by the model.

The model is then trained until convergence, such that post convergence it can perform in-context learning. That is, the model can learn to predict the correct target for a new input based on previously seen input-target pairs – provided in context. Critically, in-context learning is implemented by the model purely via its internal activations, without any additional weight updates after training. Previous work has demonstrated that this form of explicit in-context learning algorithms approximates the Bayes-optimal learning algorithm on the distribution of tasks $p(x_{1:T}, y_{1:T})$ encountered during training[12]. This key result enables us to make links between in-context learning displayed by our models and rational analysis[69].

The base neural network in our in-context learning models was the transformer-based decoder architecture[15] with a causal attention mask, as done previously[10,18,70]. The network settings were chosen based on a hyper-parameter search and were different for each domain (see SI for details), but all models irrespective of domain used positional encoding based on sine and cosine functions of different frequencies[15]. For training, a batch of tasks is sampled from $p(x_{1:T}, y_{1:T})$ in each episode and the model predicts the target for the given input conditioned on all preceding input-target pairs. After which, model parameters are updated based on the objective mentioned in Equation 4 using a schedule free optimizer[71] with the learning rate set to $3e-4$. We provide additional details about the model architecture and training procedure in the SI.

Baseline models

We chose several domain-specific cognitive models and compared them with ERMI; see SI for full details. For function learning, we compared against a meta-learned inference (MI) model trained on functions drawn from a hand-crafted prior distribution over kernels. Following Lucas et al.[35], the prior probabilities for positive linear, negative linear, quadratic, and radial basis kernels were set proportional to 8, 1, 0.1, and 0.01, respectively.

We considered six models for the domain of category learning, namely, the rational model of categorization (RMC;[43]); a meta-learned inference (MI) model trained on synthetically generated problems with linear decision boundary; a meta-learned

inference model trained on synthetically generated tasks with non-linear decision boundary (PFN;[72]); the generalized context model (GCM;[45]), a prototype model (PM;[44]), and a rule-based learning model (Rule;[34]).

Four models were considered for the decision making task. First, a meta-learned inference model trained on synthetic decision making problems sampled from the true generative model used in the experiment (MI;[49]). Second, a single-cue decision maker (SC;[49]). Third, equal weighting decision maker (EW;[49]). Fourth, a feedforward neural network (NN;[49]).

Model simulations

In this section, we provide details of how model simulations were performed for the different experiments reported in this study.

Function learning

Learning difficulty and speed. To generate the learning difficulty curves shown in Figure 2C-D, we first sampled functions, y, from linear positive, linear negative, exponential, quadratic, and periodic families, for different values of input, x, ranging between 0 and 1. For the linear functions, we used the functional form $y = mx + c$, where we sampled slope and intercept terms from uniform distribution between -1 and 1. For the exponential functions, we used $y = a * e^{bx+c} + d$, where the terms were all sampled from uniform distribution between -1 and 1. Quadratic functions used the following parameterization: $y = w^2 + c$, with values for parameters sampled from uniform distribution between -1 and 1. We used the functional form: $y = w * sin(2\pi x - \phi) + c$ for periodic function with amplitude, frequency, phase and offset sampled from uniform distribution between -1 and 1. All values were chosen such that final values are in the range between -1 to 1. We obtained the targets for each input value auto-regressively, conditioned on previous inputs and targets. We run this simulations 100,000 times for both ERMI and MI, and report the mean over trials for both models.

Interpolation and Extrapolation. We considered the same exact linear functions with fixed offset as used in the original Kwantes et al.[27] study. We only additionally normalized the input and target to be between -1 and 1, such that it matches the range of inputs taken by ERMI during training. We extracted ERMI's and MI's predictions auto-regressively, conditioning on previously observed input-target values.

Category learning

Learning difficulty. To run simulations of the Shepard study[37] on ERMI and MI, the geometric inputs used in the original study were converted into binary coded vectors taking values along the three input feature dimensions. The value assignment for a input feature was randomized in every run, the order of presentation of the input was also randomized, and the number of presentations of a input per block was matched to the original study.

Learning strategy. The 616 choices made by ERMI and MI were divided into 11 blocks of 56 trials each. The choices were obtained from the model by simulating them on a numerically abstracted version of the task, similar to the learning difficulty mentioned above. The simulations were run for a total of 50 runs using the softmax temperature term fitted to participants in the Devraj et al. 2021[41] study. We then fit prototype-model (PM) and exemplar-based model (GCM) onto the choices of humans and models to see if they are better explained by prototype or exemplar-based strategy. To fit their parameters, we minimize the sum of squared errors (SSE) between observed and predicted probabilities for each participant for a given block following the original study's approach:

$$SSE = \sum_{t=1}^{14} (p(y_t = 1|x_t) - \hat{p}_{1,x_t})^2 \tag{5}$$

where $p(y = 1|x_t)$ is the predicted probability from the model – either GCM or PM – that input x_t belongs to category 1 based on an entire trial segment (56 trials) of data, and $\hat{p}_{1,x_t}$ is the proportion of trials in the trial segment (out of those in which input x_t was seen) in which the participant or model categorized input x_t to category 1. We used SciPy's Sequential Least Squares Programming (SLSQP) method to obtain the best fitting parameter for the two models as in the original study[41]. We then compared the SSE computed using the best-fitting parameters between the two models as shown in Figure 3.

Generalization. We simulated ERMI and MI on the Johansen et al. study[39] for inverse temperature values, from zero to one in steps of 0.1, for a total of 544 runs. The models interacted with each of the nine training inputs 32 times, with the ordering of the inputs shuffled between runs. Predictions for the transfer inputs were derived by concatenating them – one at a time – at the end of 32 training blocks in every run. By doing so, we were able to derive the model's prediction for each unseen input around 77 times. In Figure 3, we reported average choice probabilities for the models using the inverse temperature value that minimized the pair-wise Euclidean distance between the human and model's choice probabilities.

Decision making

We evaluated ERMI and MI model on paired comparison tasks following the same generative model used as the original study[49]. ERMI and MI took values for the four attributes for both options along with the correct target from the previous trial as input at

each step. They then predicted one of the two option on the current step. The simulation was performed for the same number of trials and blocks as in the original study.

Model fitting and comparison

Parameters for models considered in this work were fit to the data using maximum likelihood estimation. The exact model parameters fitted for each model and their implementation details are discussed in the SI.

After fitting the models, we performed a Bayesian model comparison, with goodness-of-fit to human choices measured based on posterior model frequency[73]. The posterior model frequency measures how often a given model offers the best explanation in the population. We computed it using a Python implementation of the Variational Bayesian Analysis (VBA) toolbox[74]; see SI for additional details.

Data, code and usage of GenAI tools

Data, code, and analysis scripts are available at akjagadish/meta-learning-ecological-priors-from-llms Parts of the text were refined with the help of generative AI tools, such as ChatGPT and Writefull, which provided suggestions for rewording, paraphrasing, and restructuring. All outputs were carefully reviewed, and edited before usage.

Author contributions statement

A.K.J., M.B., and E.S. conceived the study and developed the methodology and theoretical framework; J.C. and M.T. contributed to refining the theoretical framework; A.K.J. designed and conducted the experiments with contribution from M.B.; A.K.J. collected and preprocessed the data and generated the figures; A.K.J. wrote the original draft of the manuscript; A.K.J., J.C., M.T., E.S., and M.B. reviewed and edited the manuscript; E.S. acquired funding.

Acknowledgements

This work was supported by the Max Planck Society, Helmholtz Center, Volkswagen Foundation, Princeton University, and Deutsche Forschungsgemeinschaft (DFG) under the German Excellence Strategy - EXC 2064 / 1 - 390727645.

References

1. Darwin, C. *On the Origin of Species* (John Murray, London, 1859).
2. Brunswik, E. *Perception and the Representative Design of Psychological Experiments* (University of California Press, Berkeley, 1956).
3. Simon, H. A. Rational choice and the structure of the environment. *Psychol. Rev.* **63**, 129–138, DOI: 10.1037/h0042769 (1956).
4. Gigerenzer, G., Todd, P. M. & the ABC Research Group. *Simple Heuristics That Make Us Smart* (Oxford University Press, New York, 1999).
5. Anderson, J. R. *The Adaptive Character of Thought* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1990).
6. Goldstein, D. G. & Gigerenzer, G. Models of ecological rationality: The recognition heuristic. *Psychol. Rev.* **109**, 75–90, DOI: 10.1037/0033-295X.109.1.75 (2002).
7. Brown, T. B. *et al.* Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **33**, 1877–1901 (2020).
8. Borisov, V., Seßler, K., Leemann, T., Pawelczyk, M. & Kasneci, G. Language models are realistic tabular data generators. *arXiv preprint arXiv:2210.06280* **arXiv** (2022).
9. Zhu, J.-Q. & Griffiths, T. L. Eliciting the priors of large language models using iterated in-context learning. *arXiv preprint arXiv:2406.01860* **arXiv** (2024).
10. Jagadish, A. K., Coda-Forno, J., Thalmann, M., Schulz, E. & Binz, M. Human-like category learning by injecting ecological priors from large language models into neural networks. In *Forty-first International Conference on Machine Learning* (2024).
11. Marewski, J. N., Gaissmaier, W. & Gigerenzer, G. Good judgments do not require complex cognition. *Cogn. Process.* **10**, 117–128, DOI: 10.1007/s10339-009-0264-y (2009).
12. Ortega, P. A. *et al.* Meta-learning of sequential strategies. *arXiv preprint arXiv:1905.03030* **arXiv** (2019).
13. Binz, M. *et al.* Meta-learned models of cognition. *Behav. Brain Sci.* **47**, e147 (2024).

14. Bischl, B. *et al.* Openml benchmarking suites. *arXiv:1708.03731v2 [stat.ML]* **arXiv** (2019).

15. Vaswani, A. *et al.* Attention is all you need. *Adv. neural information processing systems* **30** (2017).

16. Hochreiter, S., Younger, A. S. & Conwell, P. R. Learning to learn using gradient descent. In *Artificial Neural Networks—ICANN 2001: International Conference Vienna, Austria, August 21–25, 2001 Proceedings 11*, 87–94 (Springer, 2001).

17. Wang, J. X. *et al.* Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763* **arXiv** (2016).

18. Lake, B. M. & Baroni, M. Human-like systematic generalization through a meta-learning neural network. *Nature* **Nature**, 1–7 (2023).

19. Carroll, J. D. Functional learning: The learning of continuous functional mappings relating stimulus and response continua. *ETS Res. Bull. Ser.* **1963**, i–144 (1963).

20. Koh, K. & Meyer, D. E. Function learning: Induction of continuous stimulus–response relations. *J. Exp. Psychol. Learn. Mem. Cogn.* **17**, 811–836, DOI: 10.1037/0278-7393.17.5.811 (1991).

21. Brehmer, B. Hypothesis testing, probability learning, and a simple rule. *J. Exp. Psychol.* **102**, 887–891, DOI: 10.1037/h0036211 (1974).

22. DeLosh, E. L., Busemeyer, J. R. & McDaniel, M. A. Extrapolation: the sine qua non for abstraction in function learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **23**, 968 (1997).

23. Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M. & Gershman, S. J. Compositional inductive biases in function learning. *Cogn. psychology* **99**, 44–79 (2017).

24. Schulz, E., Tenenbaum, J. B., Reshef, D. N., Speekenbrink, M. & Gershman, S. J. Assessing the perceived predictability of functions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 37 (2015).

25. Lichtenberg, J. M. & Şimşek, Ö. Simple regression models. In *Imperfect decision makers: Admitting real-world rationality*, 13–25 (PMLR, 2017).

26. Busemeyer, J. R., Byun, E., Delosh, E. L. & McDaniel, M. A. Learning functional relations based on experience with input–output pairs by humans and artificial neural networks. In *Knowledge concepts and categories*, 405–437 (Psychology Press, 2013).

27. Kwantes, P. J. & Neal, A. Why people underestimate y when extrapolating in linear functions. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 1019 (2006).

28. Kalish, M. L., Lewandowsky, S. & Kruschke, J. K. Population of linear experts: knowledge partitioning and function learning. *Psychol. review* **111**, 1072 (2004).

29. Little, D. R., Shiffrin, R. M. & Laham, S. M. Function estimation: Quantifying individual differences of hand-drawn functions. *Mem. & Cogn.* **Springer**, 1–20 (2024).

30. Brehmer, B. Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organ. Behav. Hum. Perform.* **11**, 1–27 (1974).

31. Brehmer, B. Single-cue probability learning as a function of the sign and magnitude of the correlation between cue and criterion. *Organ. Behav. Hum. Perform.* **9**, 377–395 (1973).

32. Brehmer, B., Kuylenstierna, J. & Liljergren, J.-E. Effects of function form and cue validity on the subjects' hypotheses in probabilistic inference tasks. *Organ. Behav. Hum. Perform.* **11**, 338–354 (1974).

33. Byun, E. *Interaction between prior knowledge and type of nonlinear relationship on function learning*. Ph.D. thesis, Purdue University (1995).

34. Mcdaniel, M. A. & Busemeyer, J. R. The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models. *Psychon. bulletin & review* **12**, 24–42 (2005).

35. Lucas, C. G., Griffiths, T. L., Williams, J. J. & Kalish, M. L. A rational model of function learning. *Psychon. bulletin & review* **22**, 1193–1215 (2015).

36. Ashby, F. G. & Maddox, W. T. Human Category Learning. *Annu. Rev. Psychol.* **56**, 149–178, DOI: 10.1146/annurev.psych.56.091103.070217 (2005).

37. Shepard, R. N., Hovland, C. I. & Jenkins, H. M. Learning and memorization of classifications. *Psychol. Monogr. Gen. Appl.* **75**, 1–42, DOI: 10.1037/h0093825 (1961).

38. Smith, J. D. & Minda, J. P. Prototypes in the mist: The early epochs of category learning. *J. Exp. Psychol. Learn. memory, cognition* **24**, 1411 (1998).

39. Johansen, M. K. & Palmeri, T. J. Are there representational shifts during category learning? *Cogn. Psychol.* **45**, 482–553, DOI: 10.1016/s0010-0285(02)00505-4 (2002).

40. Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C. & Glauthier, P. Comparing models of rule-based classification learning: a replication and extension of shepard, hovland, and jenkins (1961). *Mem. Cogn.* **22**, 352–369, DOI: 10.3758/bf03200862 (1994).

41. Devraj, A., Zhang, Q. & Griffiths, T. The dynamics of exemplar and prototype representations depend on environmental statistics. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 43 (2021).

42. Badham, S. P., Sanborn, A. N. & Maylor, E. A. Deficits in category learning in older adults: Rule-based versus clustering accounts. *Psychol. Aging* **32**, 473–488, DOI: 10.1037/pag0000183 (2017).

43. Anderson, J. R. The adaptive nature of human categorization. *Psychol. Rev.* **98**, 409–429, DOI: 10.1037/0033-295X.98.3.409 (1991).

44. Homa, D. & Cultice, J. C. Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 83 (1984).

45. Nosofsky, R. M. Attention, similarity, and the identification–categorization relationship. *J. experimental psychology: Gen.* **115**, 39 (1986).

46. Müller, S., Hollmann, N., Arango, S. P., Grabocka, J. & Hutter, F. Transformers can do bayesian inference. *arXiv preprint arXiv:2112.10510* **arXiv** (2021).

47. Samuels, R., Stich, S. & Bishop, M. Ending the rationality wars. *Collect. Pap. Vol. 2: Knowledge, Ration. Morality, 1978-2010* **2**, 191 (2012).

48. Camerer, C., Loewenstein, G. & Prelec, D. Neuroeconomics: How neuroscience can inform economics. *J. economic Lit.* **43**, 9–64 (2005).

49. Binz, M., Gershman, S. J., Schulz, E. & Endres, D. Heuristics from bounded meta-learned inference. *Psychol. review* **Psychological review** (2022a).

50. Geisler, W. S. Sequential ideal-observer analysis of visual discriminations. *Psychol. review* **96**, 267 (1989).

51. Gigerenzer, G. & Gaissmaier, W. Heuristic decision making. *Annu. review psychology* **62**, 451–482 (2011).

52. Chater, N. & Vitányi, P. Simplicity: a unifying principle in cognitive science? *Trends cognitive sciences* **7**, 19–22 (2003).

53. Todd, P. M. & Gigerenzer, G. Environments that make us smart: Ecological rationality. *Curr. directions psychological science* **16**, 167–171 (2007).

54. Martignon, L. & Hoffrage, U. Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory Decis.* **52**, 29–71 (2002).

55. Goldstein, D. G. & Gigerenzer, G. The recognition heuristic: How ignorance makes us smart. In *Simple heuristics that make us smart*, 37–58 (Oxford University Press, 1999).

56. Farrell, H., Gopnik, A., Shalizi, C. & Evans, J. Large ai models are cultural and social technologies. *Science* **387**, 1153–1156 (2025).

57. Tenenbaum. Joshua Tenenbaum's homepage. http://web.mit.edu/cocosci/josh.html (2021). [Online; accessed 9-November-2021].

58. Todd, P. M. & Brighton, H. Building the theory of ecological rationality. *Minds Mach.* **26**, 9–30, DOI: 10.1007/s11023-015-9371-0 (2016).

59. Jagadish, A. K., Binz, M., Saanum, T., Wang, J. X. & Schulz, E. Zero-shot compositional reinforcement learning in humans. *PsyArXiv preprint PsyArXiv:ymve5* **PsyArXiv** (2023).

60. Jagadish, A. K., Coda-Forno, J., Thalmann, M., Binz, M. & Schulz, E. Bounded ecologically rational meta-learned inference explains human category learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 47 (2025).

61. Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D. & Griffiths, T. L. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* **372**, 1209–1214 (2021).

62. Ji-An, L., Benna, M. K. & Mattar, M. G. Automatic discovery of cognitive strategies with tiny recurrent neural networks. *bioRxiv* **bioRxiv**, 2023–04 (2023).

63. Miller, K., Eckstein, M., Botvinick, M. & Kurth-Nelson, Z. Cognitive model discovery via disentangled rnns. *Adv. Neural Inf. Process. Syst.* **36**, 61377–61394 (2023).

64. Bauer, J. *et al.* Human-timescale adaptation in an open-ended task space. *PMLR* **PMLR**, 1887–1935 (2023).

65. Anthropic, P. B. C. Claude 2. https://www.anthropic.com/index/claude-2 (2023). Accessed: 2024-1-15.

66. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

67. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17**, 261–272, DOI: 10.1038/s41592-019-0686-2 (2020).

68. Chan, S. *et al.* Data distributional properties drive emergent in-context learning in transformers. *Adv. Neural Inf. Process. Syst.* **35**, 18878–18891 (2022).

69. Anderson, J. R. Is human cognition adaptive? *Behav. brain sciences* **14**, 471–485 (1991).

70. Schubert, J. A., Jagadish, A. K., Binz, M. & Schulz, E. In-context learning agents are asymmetric belief updaters. In *International Conference on Machine Learning*, 43928–43946 (PMLR, 2024).

71. Defazio, A. *et al.* The road less scheduled. *Adv. Neural Inf. Process. Syst.* **37**, 9974–10007 (2024).

72. Müller, S., Hollmann, N., Arango, S. P., Grabocka, J. & Hutter, F. Transformers can do bayesian inference. In *International Conference on Learning Representations* (2022).

73. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies—revisited. *Neuroimage* **84**, 971–985 (2014).

74. Daunizeau, J., Adam, V. & Rigoux, L. Vba: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS computational biology* **10**, e1003441 (2014).

Supplementary Information for Meta-learning ecological priors from large language models explains human learning and decision making

Akshay K. Jagadish[1,2,3,4,*], **Mirko Thalmann**[2], **Julian Coda-Forno**[2], **Marcel Binz**[2,+], **and Eric Schulz**[2,+]

[1]MPRG Computational Principles of Intelligence, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2]Institute for Human-Centered AI, Helmholtz Computational Health Center, Munich, Germany
[3]Eberhard Karls University of Tübingen, Tübingen, Germany
[4]Princeton University, Princeton, United States of America
[*]akshaykjagadish@gmail.com
[+]these authors contributed equally to this work

Function Learning

LLM-generated tasks

The exact prompts and data generation pipeline for function learning are discussed in the Methods section of the main text.

Parsing synthesized task features and labels: We queried CLAUDE-V2 to generate feature names in the format: FEATURE DIMENSION 1, FEATURE DIMENSION 2, ...; see Methods in the main text for exact prompts. To extract these, we used a series of regex patterns, such as `([A-Za-z&]+),([A-Za-z&]+)` and its higher-arity extensions, designed to capture up to five comma-separated alphanumeric feature names (including symbols like "&"). These patterns allowed us to reliably extract structured feature descriptions across tasks. The parsed feature names were stored in a dataframe for subsequent task construction and evaluation.

Qualitative analysis of synthesized task features and labels: We show the counts for the top-50 most frequently occurring names for (a) inputs and (b) targets in Figure S1. We found that LLM tends to produce input-target pairs belonging to everyday topics such as education (practice time versus skill), health (calories burned versus weight change), agriculture (rainfall versus crop yield), etc.

Parsing generated task data points: CLAUDE-V2 was prompted to generate datapoints in the format: - FEATURE VALUE 1, FEATURE VALUE 2, ..., FEATURE VALUE N, TARGET VALUE; see Methods in the main text for exact prompts. To extract numerical values from these responses, we constructed regex expressions of the form `([\d.]+),` repeated for each feature dimension, followed by `([\d.]+)` for the target value. This pattern reliably captured sequences of decimal numbers across varying dimensionalities. The extracted values were stored in a dataframe, serving as a structured dataset for training and evaluating meta-learned function approximators.

Data processing: We filter out all tasks containing more than 20 data points to ensure consistent task lengths and evaluation settings. We randomly shuffled the trial order within each task. We resampled the trials with replacement to match the target task duration, enabling evaluation on experiments with longer trial horizons without performance degradation. All feature dimensions were independently normalized to lie within $[-\text{scale}, \text{scale}]$ using a Min-Max normalization scheme, where $\text{scale} \in [0.1, 0.5]$ was fixed or randomly sampled. LLM-generated tasks can sometimes be of varying lengths, and in the case that the task length was shorter, they were padded with zeros to match the longest task in the batch. The maximum steps or number of trials for the experiment we considered was 25 trials. The batch size was fixed to 64 unless otherwise specified.

Models fit to LLM-generated data: We considered models from four well-studied function families, namely, linear, exponential, quadratic, and sinusoidal, as mentioned in the Methods. For the linear function, we chose the instantiation $y = a*x+b$, with initial parameters set to 1 and 0 for the slope and offset terms, respectively. For quadratic, we chose $y = a*x^2+c$, with initial parameters for slope and offset set to 1 and 0 respectively. We chose $y = a*exp(b*x)+d$ for the exponential family, with initial parameters for a set to the mean difference between the maximum and minimum of the target values, b set to 1, and offset term set to the minimum of the target values. We chose $y = a*sin(b*x)+d$ for the sinusoidal family with initial parameters a set to the mean difference between the maximum and minimum of the target values, b set to $2*\pi$, and offset

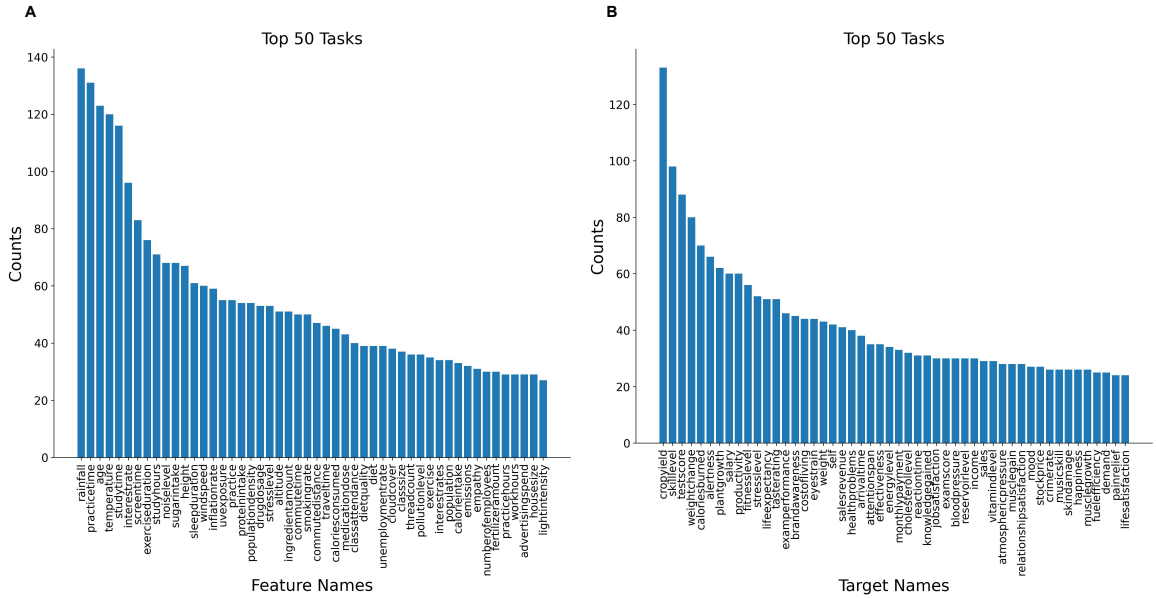

Figure S1. Frequency of input and target labels in CLAUDE-V2 synthesized function learning tasks: Counts for the top-50 most frequently occurring (a) input and (b) target labels computed over 9991 LLM-generated function learning tasks. These distributions confirm that the LLM-generated tasks capture real-world functional relationships.

term set to the mean of the target values. We fit the parameters of these models to LLM-generated functions using the curve fit function from the SCIPY optimization library[1].

Human studies

Kwantes and Neal 2006[2]. In this study, 14 participants had to learn to predict values along the y-axis for different values on the x-axis, with samples drawn from a linear function $y = 2.2x + 30$. Before test phase, they were trained on 20 samples on the x-axis drawn such that their values on the y-axis were always in the range between 30 and 70. The samples were fixed but their order used for training was randomized per participant and session. In each trial, participants made their prediction by entering their estimate as numbers and locking in their answer by clicking on a button labeled “submit your answer". After locking in, feedback was provided regarding their performance (in terms of accuracy score out of 100). Once training was complete, participants were shown 45 samples in the range from 0 to 100 and asked to enter their estimates. The presentation of the 45 samples were blocked into three sets of 15: low (0-30), medium (30-70), and high (70-100) range. The order in which the blocks were presented and the order of samples within them was randomized for each participant.

Little et al. 2024[3]. This study was conducted on 177 participants. The particular experiment we consider, called function estimate test, was included as part of larger paper-based questionnaire. In this experiment, participants were presented 24 scatter plots, each depicting data from a different fictional scientific experiment, on a piece of paper, with two 7.5 cm by 7.5 cm graphs in each page with 4 cm gap between them. They were then instructed to draw the true underlying causal function for the data points in the graph. The data points could be presented in either large (zoomed in version), where the data points covered the entire figure, or small (zoomed out version), where it covered 40 percent of the total area, scale. The relative position of the points in the small- and large-scale sets was kept identical. Three functions were used to generate data for the scatter plots, namely, linear, quadratic, or cubic polynomial functions. A small amount of Gaussian noise was added as jitter in all graphs. The data points and the drawn functions used for model fitting were extracted from scans of the physical document using a software program called Data Thief[4]. After extraction, the data was down-sampled to include 40 evenly spaced data points in the range of the x-axis and with all points scaled to be between -1 and 1. We used the data from the following GitHub repository.

Hand-crafted tasks

Functional priors from rational model of function learning[5] used for training MI model: We generate 10,000 synthetic regression tasks for function learning using a mixture of kernels adapted from the study by Lucas et al.[5]. Each task involved a one-dimensional input sampled from a uniform grid of 20 points in the interval $[-1, 1]$. The target output was computed by sampling a kernel type from a hand-crafted prior: favoring positive linear (probability 0.8), followed by negative linear (0.1), quadratic (0.01), and radial basis (0.001) functions and applying the corresponding transformation to the input. Parameters for each kernel (e.g. weights, intercepts, distances) were drawn from a gamma distribution with shape 1.001 and scale 1.0. A small amount of Gaussian noise was added to the target. All inputs and targets were dynamically scaled to lie in $[-\text{scale}, \text{scale}]$, where the scale is sampled from a uniform distribution in the range $[0.1, 0.5]$ in each training batch.

Model architecture, and training

Each trial in a function learning task consisted of an input vector concatenated with the previous target value, and these were embedded into a 64-dimensional space. Positional encoding was applied using sine and cosine functions of varying frequencies, following Vaswani et al.[6]. A causal attention mask ensured that predictions at each time step were conditioned only on past inputs and targets. These masked sequences were processed using a Transformer decoder composed of six layers, with 64-dimensional embeddings, eight attention heads, and 256 hidden units in the feedforward layers. The decoder outputs were passed through two independent linear projections to produce the mean and standard deviation of a normal distribution. The negative log-likelihood (NLL) was computed over all targets in a given batch, and minimizing it served as a loss function for training the network parameters. The model parameters were updated using the ScheduleFree optimizer[7] with a baseline learning rate of 3×10^{-4}. Each model was trained for 250.000 episodes, with periodic evaluation on held-out tasks to monitor generalization performance.

Model fitting and comparison

We did not fit any model parameters to human data in both ERMI and MI. We computed the response from these models by querying it on new inputs while being conditioned on the input-target pairs participant observed before drawing the functions. For model comparison, we report the mean-squared error between the participant's actual response, sampled from the functions they drew through the data points displayed to them, and model predicted target for the same input.

Additional results

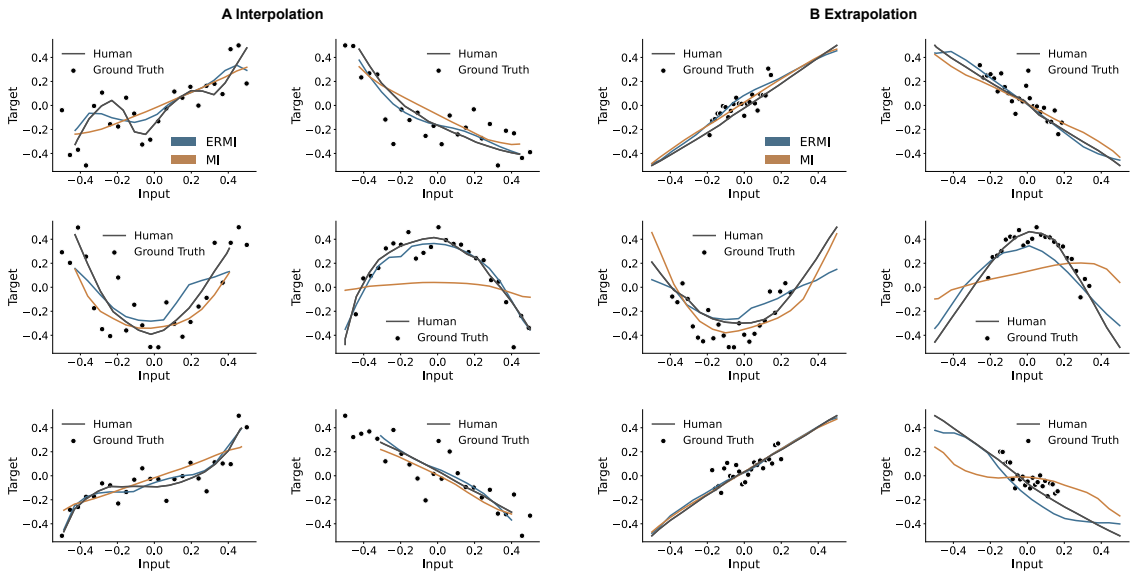

Figure S2. Predictions derived from ERMI and MI for function families used in the Little et al. study[3] for both interpolation (zoomed in; A) and extrapolation (zoomed out; B) condition. The function families considered were linear (top row), quadratic (middle row) and cubic polynomial (bottom row); see Human studies section in Methods for details.

Category Learning

LLM-generated tasks

Prompts: We used the following prompt to synthesize feature names and category labels for the category learning task.

Synthesize feature names and category labels

I am a psychologist who wants to run a category learning experiment. In a category learning experiment, there are many different three-dimensional stimuli, each of which belongs to one of two possible real-world categories.

Please generate names for three stimulus feature dimensions and two corresponding categories for 250 different category learning experiments:

In the second stage, we prompted the LLM to generate data points for the synthesized features and the category label. Below is the prompt, for a category learning where the synthesized input features were sodium, fat, and protein, and categories are healthy or unhealthy:

Generate category learning tasks

I am a psychologist who wants to run a category learning experiment. For a category learning experiment, I need a list of stimuli and their category labels. Each stimulus is characterized by three distinct features: **sodium**, **fat**, and **protein**. These features can take only numerical values. The category label can take the values **healthy** or **unhealthy** and should be predictable from the feature values of the stimulus.

Please generate a list of 100 stimuli with their feature values and their corresponding category labels using the following template for each row:

– feature value 1, feature value 2, feature value 3,
category label

Parsing synthesized task features and labels: We prompted CLAUDE-V2 to generate task features and labels in the format: FEATURE DIMENSION 1, FEATURE DIMENSION 2, ..., FEATURE DIMENSION N, CATEGORY LABEL 1, CATEGORY LABEL 2. We extracted relevant entries using the regex pattern `\d+. (.+?)\n`, which captures text following a numbered bullet point up to the first newline. The resulting string was split at the commas to separate feature names from category labels. All parsed information was stored in a dataframe for downstream use.

Qualitative analysis of synthesized task features and labels: We show the counts for the top-50 most frequently occurring input feature names in Figure S3 and category names in Figure S4 for the 23421, 20690, and 13693 category learning tasks generated with three (a), four (b) and six-dimensional input features, respectively. When it comes to input feature names, we found that the LLM tends to produce features belonging to topics such as musicality (like rhythm, melody, lyrics, tempo, vocals), food (like aroma, texture, crust, diet, protein), etc. With regard to category names, there were also many related to music (for example, classical, pop, jazz, rock), but also vehicles (like trucks, SUVs, sedans), technology (laptops, desktops, iPads), etc.

Parsing generated task data points: To generate data points for each task, we queried CLAUDE-V2 using the format: - FEATURE VALUE 1, FEATURE VALUE 2, ..., FEATURE VALUE N, CATEGORY LABEL. The model reliably followed this format. To parse the resulting output, we used a suite of regex patterns designed to handle diverse data formats, including numeric values (with or without decimals), alphanumeric labels, hyphens, and various delimiters. Table S1 lists all the regex patterns employed. These enabled us to successfully parse 95% of the generated tasks. The parsed values were stored in a dataframe, forming an offline task repository to train the ecologically rational meta-learned inference model.

Data pre-processing: We filter out all tasks with more than two unique category labels and then binarize the category labels, which are originally strings, to make them consistent across tasks. The assignment of category labels, that is either '0' or '1', within a category learning task was randomized during batch creation. This ensures that there can be no unintended correlations between the inputs seen during training and the labels (across all training data each input vector is assigned half of the time to label '0' and half of the time to label '1'). We also normalized each feature independently using a min-max normalization scheme such that values taken by any feature lie always between zero and one. Both the task features and data points were shuffled while generating tasks. Note that the tasks generated by LLMs are typically of different lengths. Whenever the sampled

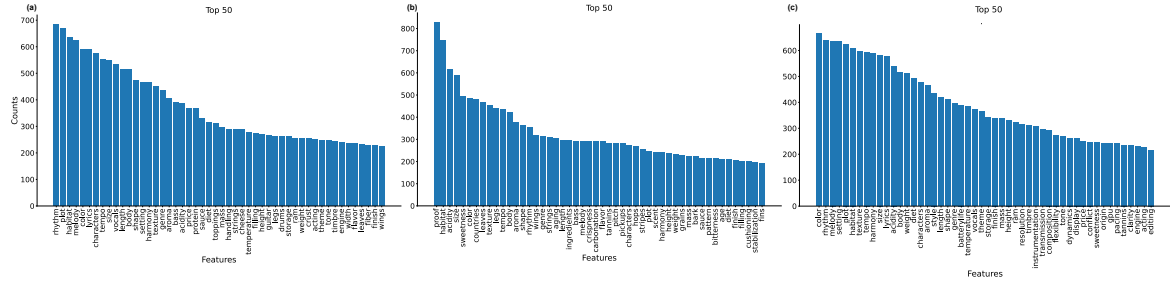

Figure S3. Frequency of input feature names in CLAUDE-V2 synthesized category learning tasks: Counts for the top-50 most frequently occurring input features in the 23421, 20690, and 13693 category learning tasks generated for three (a), four (b), and six-dimensional features respectively. These distributions confirm that the LLM-generated real-world features relevant for category learning.

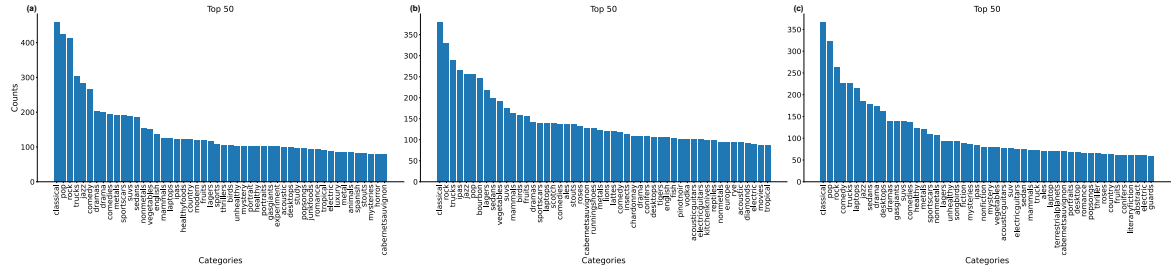

Figure S4. Frequency of different category names in CLAUDE-V2 synthesized category learning tasks: Counts for the top-50 most frequently occurring category names in the 23421, 20690, and 13693 category learning tasks generated for three (a), four (b), and six-dimensional features respectively. These distributions confirm that the LLM-generated category names prevalent in the real-world.

tasks are of variable lengths, they are padded with zeros to match the length of the longest task sample within the batch. We additionally also sampled LLM-generated data points with replacement to match the length of the experimental task used in the Devraj et al. 2021[8] and Johansen et al. 2002[9] studies. We resorted to this strategy as the LLM-generated tasks had a maximum of about 200 data points per task and by resampling, we can evaluate the model on experiments with larger horizons without any drop in performance. The batch size was set to 64 for three- and four-dimensional inputs and to 32 for six-dimensional inputs and it operated under a maximum steps regime of 400, 300, and 650 for three, four, and six-dimensional tasks respectively.

Human studies

Nosofsky et al. 1994[10]. In their replication of the Shepard et al. 1961[11] study, Nosofsky and colleagues conducted the study on 120 participants. The authors used geometric inputs that varied in shape (squares or triangles), interior line type (solid or dotted), and size (large or small). In total, 40 participants performed each of the six category structures, considered in Shepard et al. 1961[11]. The participants were informed that the rules for each problem were independent. Following the same methodology as Shepard et al., the learning process involved classifying inputs into two categories and receiving feedback. This process was repeated over several blocks (containing up to 16 trials) with randomized input order in each block. Learning in the task was measured until participants achieved a no-error streak in four consecutive sub-blocks of eight trials or reached a maximum of 400 trials. In tasks belonging to TYPE 1, inputs were assigned to a category depending on the values they take along one of the three dimensions, whereas in TYPE 2 tasks, inputs were assigned to a category by applying the exclusive-or rule along two relevant dimensions. Category assignment in tasks belonging to TYPE 3, TYPE 4, and TYPE 5 used a unidimensional rule-plus-exception structure with some inputs grouped in the central region and some in the periphery. Lastly, TYPE 6 tasks required considering feature values along all dimensions, and they require the memorization of every item and its associated category to solve them correctly. For the illustration of category structures for the six types, please refer to Figure 1 in Nosofsky et al. study[10].

Table S1. Regular expression patterns used for parsing the data points generated for category learning tasks by CLAUDE-V2

INDEX	REGULAR EXPRESSION
1	`([\d.]+),([\d.]+),([\d.]+),([\w]+)`
2	`([\w\-]+),([\w\-]+),([\w\-]+),([\w]+)`
3	`([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+)`
4	`([^,]+),([^,]+),([^,]+),([^,]+)`
5	`([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+)`
6	`(?:.*?:)?([^,-]+),([^,-]+),([^,-]+),([^,-]+)`
7	`([^,-]+),([^,-]+),([^,-]+),([^,-]+)`
8	`r'^(\d+):([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\w]+)'`
9	`r'^(\d+):([\w\-]+),([\w\-]+),([\w\-]+),([\w\-]+),([\w]+)'`
10	`r'^(\d+):([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+),([-\w\d,.]+)'`
11	`r'^(\d+):([^,]+),([^,]+),([^,]+),([^,]+),([^,]+)'`
12	`r'^(\d+):([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+)'`
13	`r'^(\d+):(?:.*?:)?([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)'`
14	`r'^(\d+):([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)'`
15	`^(\d+):([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\d.]+),([\w]+)`
16	`^(\d+):([\w-]+),([\w-]+),([\w-]+),([\w-]+),([\w-]+),([\w-]+),([\w]+)`
17	`(\d+):([^,]+),([^,]+),([^,]+),([^,]+),([^,]+),([^,]+),([^,]+)`
18	`(\d+):([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+),([^,\n]+)`
19	`(\d+):(?:.*?:)?([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)`
20	`(\d+):([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+),([^,-]+)`

Badham et al. 2017[12]. In this study, the authors partially replicated the original Shepard et al. 1961[11] study by running it on 96 adults aged between 18 to 87 years. As inputs, they used eight geometric shapes varying in size (large or small), shape (square or triangle), and color (black or white) shown on a mid-gray background. The order of inputs and their category assignment were randomized. Unlike the original study, the authors only considered the first four types of category structures but with the advantage that all participants performed all four types. Participants performed each task type for a total of six blocks with each block containing 16 trials (resulting in a total of 96 trials) or until they reached a criterion of perfect performance in two consecutive blocks.

Smith et al. 1998[13]. The study was run on 32 participants, where each participant was presented 14 different six-dimensional inputs, with each input mapping to a six-letter nonsensical word such as gafuzi, kafitdo, nivety, wysero, etc (see Appendix A of[13] for all words). For modeling, we represented each input as a six-digit binary string, where each digit and position corresponds to a specific letter. For instance, assuming the input 'gafuzi' corresponds to the binary code '000000', 'gyfuzi' corresponds to '010000', and so on. The inputs were assigned to categories such that input '000000' corresponds to category 1 and input '11111' corresponds to category 2. In this work, we only considered data from the non-linearly separable (NLS) category structure from Experiment 2. In this category structure, each category consisted of six inputs with five features in common with the prototype, and one input with five features in common with the opposing prototype. For instance, if category 1 contained seven inputs as follows: [000000, 100000, 010000, 001000, 000010, 000001, 111101]. The remaining seven inputs belonged to category 2 [111111, 011111, 101111, 110111, 111011, 111110, 000100]. Participants had to categorize a input into one of these two categories and had unlimited time to make their choices. After making their choice, they were told if it was a correct decision or not. Participants completed a total of 560 trials over 10 blocks of 56 trials each. In each block, participants saw each input four times.

Devraj et al. 2021[8]. Devraj and colleagues replicated a study of Smith et al. 1998[13] and collected data from 60 participants. Participants were recruited from the 18-23 age range and English-speaking population using Prolific. Their study involved 11 blocks and had 616 trials in total. We used the data from the following GitHub repository.

Johansen et al. 2002[9]. Johansen and colleagues conducted their categorization study on 130 participants in which they presented four-dimensional inputs with each dimension taking binary values. Each of the inputs was a computer-generated drawing of a rocket that varied along four binary-valued dimensions: The shape of the wing (triangular or rectangular), tail

(jagged or boxed), nose (staircase or half-circle), and porthole (circular or star)[9]. The authors used the same category structure as those used in previous studies[14,15]. This category structure is ill-defined in that no single feature along a dimension can be used to perfectly classify inputs. Instead, the categories have a family resemblance structure in that category 1 inputs tend to have a value of 0 along each dimension, and category 2 inputs tend to have a value of 1 along each dimension. More concretely, they assigned they five inputs [0001, 0101, 0100, 0010, 1000] to category 1 and the remaining four inputs [0011, 1001, 1110, 1111] to category 2. The inputs were presented serially with their order randomized within each block. Participants had unlimited time to make their choice and were informed whether or not it was a correct choice after each choice. Participants completed a total of 288 training trials, or 32 blocks of 9 trials each, in which they saw each input once. In addition to the training block, participants had to perform a transfer block after 2, 4, 8, 16, 24, and 32 blocks of training. In a transfer block, the eight training inputs along with eight other unseen transfer inputs were shown without corrective feedback. The encoding for transfer inputs, labeled T1 to T7, were (in order): [1011, 1010, 0111, 1101, 1100, 0110, 0000]. It is the category assigned in the transfer block, which is of major interest in this work.

Hand-crafted tasks

Bayesian logistic regression prior used for training MI model: We generated 10.000 synthetic binary classification tasks with a linear decision boundary using a Bayesian logistic regression model. To do this, we sample the input features from a normal distribution with zero mean and unit variance for a given number of data points and input dimensions. We then applied a linear transformation, followed by a sigmoid function, and rounded the result to determine the binary class for the given input. The parameters of the linear transformation are sampled from a normal distribution with zero mean and unit variance. The maximum number of data points within a task was set to 400, 650, or 300 for category learning tasks with three-, four-, and six-dimensional inputs, respectively. These values were chosen according to the length of the experiments on which these models were evaluated.

Bayesian neural network prior used for training prior-fitted network (PFN) model: We generated 10.000 synthetic binary classification tasks using a version of the Bayesian neural network (BNN) developed by Müller et al.[16]. We used normally-distributed i.i.d. input features for a given number of data points and input dimensions. We then passed the input through a BNN with two layers with tanh non-linearity and hidden dimensionality of 64. The network weights and biases were sampled from a normal distribution with a mean of zero and a standard deviation of 0.1 and subjected to an additional sparsity constraint (i.e., 20 percent of randomly chosen network weights and biases set to zero). The maximum number of data points was once again set to 400, 650, or 300 for category learning tasks with three-, four-, and six-dimensional inputs, respectively. The model output is passed through a sigmoid function to generate probability estimates, which are then rounded to determine the class for the given input.

Model architecture, and training

The task features, which contain values for the different input features and the target from the previous trial, were mapped to a 64-dimensional embedding space and positional encoded using sine and cosine functions of different frequencies as in Vaswani et al.[6]. Then a causal attention mask was generated for the inputs so that the model makes conditional predictions on all preceding data points. The inputs along the attention mask are then passed to the transformer decoder model, which has six layers, a model dimension of 64, 256 hidden units in the feed-forward network, and eight attention heads. The output of the transformer was then passed through a linear readout and sigmoid function to generate probability estimates for category 1. In practice, inference for all time steps is performed in parallel by passing a causal attention mask to the transformer decoder module in PyTorch[17]. We used binary cross-entropy (BCE) loss for a given batch of inputs and updated the model parameters using the ADAM optimizer[18] with a learning rate of 10^{-4}. We trained all our models for a total of 500.000 episodes.

Baseline models

Apart from models derived by meta-learning on hand-crafted priors, we considered four other cognitive models as baselines in the domain of category learning, as detailed below.

Rational model of categorization (RMC): The RMC is a Bayesian model of human category learning developed by Anderson et al.[19]. To derive this model, we simulated data from underlying generative model, such that it followed the data-generating distribution described in Badham et al.[12], and meta-learned on the generated data, similar to meta-learning on hand-crafted priors. The architecture and training of the model followed the protocol used for ERMI, MI and PFN. We set the free parameters for the RMC based on an earlier study[10] to the following values: $c = 0.318$, $s_P = 0.488$, and $s_L = 0.046$. However, we did not account for these parameters in our model comparisons, which could explain why the predictive performance RMC is overestimated.

Prototype-based model (PM): Over the years, many different versions of the prototype model have been produced[13,14]. We used the version from Smith et al.[13]. This model assigns a category to an observed stimulus based on the similarity distance to the prototype from each category. Specifically, the similarity distance between the stimulus and a prototype, q_k, for category k is calculated as a weighted sum of absolute differences in the dimensions of the features n, with $w_j \in [0,1]$ corresponding to the weights per feature. The weights are normalized to sum up to 1 as shown in Equation 1.

$$d_{x,q_k} = \sum_{j=1}^{n} w_j \left| x_j - q_{k,j} \right|, \tag{1}$$

The prototypes themselves can be learned or directly specified during model definition. In our case, we assume the prototypes for the two categories $\{q_1, q_2\}$ as a learnable parameter and learn them during the model fitting procedure. That is, $q_{k,j} \in [0,1.] \; \forall j = \{1,2,...n\}$ are assumed to be learnable model parameters. The similarity distance between prototypes and stimuli is converted into a psychological space using:

$$\eta_{x,q_k} = e^{-c \cdot d_{x,q_k}} \tag{2}$$

where c is a sensitivity parameter that can shrink or amplify discriminability in a psychological space. The probability of the stimulus being assigned to the category $k = 1$ was then calculated using the following.

$$P(k = 1 \mid x) = \frac{\eta_{x,q_1}}{\eta_{x,q_1} + \eta_{x,q_2}} \tag{3}$$

Furthermore, the predicted likelihood of the final model is a mixture between the predicted probability of the model and a random guess, with the guessing parameter ε controlling the mixture probabilities.

$$p(k = 1 \mid x) = (1 - \varepsilon) P(k = 1 \mid x) + \varepsilon \cdot \mathrm{K}^{-1} \tag{4}$$

where K indicates the number of categories.

Generalized context model (GCM): GCM is an exemplar-based model of human category learning developed by Nosofsky et al.[20]. The GCM assigns an observed stimulus to a category by comparing the sum of its similarity scores to all previously seen exemplars in each category, $\{C_1, C_2\}$. The raw distance between the observed stimulus and the exemplars and the similarity score were calculated based on Equations 1 and 2, respectively. The posterior probability of category membership $k = 1$ is calculated on the basis of normalized similarity scores as follows.

$$P(k = 1 \mid x) = \frac{\sum_{y \in C_1} \eta_{x,y}}{\sum_{y \in C_1} \eta_{x,y} + \sum_{y \in C_2} \eta_{x,y}} \tag{5}$$

The final likelihood of category membership is computed as a mixture between the estimate of posterior probability and a random guessing model as mentioned in Equation 4.

Rule: The rule model considered as the baseline in this work assigns a stimulus to a category based on one of the two rules, whichever better explains the choices of the participants. The first rule is based on the values taken by stimulus features along one dimension, and the second is based on the application of the conjunctive rule on pairs of features, whether a given pair of stimulus features takes on the same value. The final category membership is determined by a mixture between the predicted posterior class probabilities of the model and a random guess, as discussed in Equation 4.

Model fitting

The parameters of all models in the domain of category learning were fit to human data using maximum likelihood estimation. We explain the exact implementation details for the different model classes in the following. The complete list of the parameters fitted for each model is shown in Table S2.

MI, PFN, RMC and ERMI: For models derived using meta-learning, we fitted the inverse temperature term β within the sigmoid function, which squashes the output from the final layer of the transformer to be within $[0,1]$, to each participant. This term was set to a value of 1 during meta-learning to allow us to derive a Bayes-optimal model and was only fitted during the evaluation phase (bounded to be within $[0,10]$), with the rest of the model weights frozen. For parameter fitting, we used the differential evolution optimizer available in the SciPy optimization library[1].

GCM and PM: We fit the three parameters common to GCM and PM, namely feature weights, sensitivity, and the random guessing parameter, with feature weights bounded to lie within the range $[0,1]$ and summing to 1; sensitivity term bounded to lie within the $[0,20]$ range; and the guessing parameter bounded to be within $[0,1]$. The prototype model also required learning the prototypical stimulus for each category, which is of the same dimensionality as the input stimulus, with the feature values bounded within $[0,1]$. For parameter fitting, we used the MINIMIZE module available in the SCIPY optimization library.

Rule: We used the same procedure as above except that we learn the stimulus dimension v_i on which the rule is applied.

CLAUDE-V2: We used the same procedure as above except that only the guessing parameter, ε, is learned.

Table S2. This table provides the complete list of model parameters that were fit to human data in the domain of category learning, where β is the inverse temperature term, w_i indicates the weights for the stimulus feature dimension i, n is the number of stimulus feature dimensions, c is the sensitivity term, ε is noise term in an epsilon greedy policy, q_1 and q_2 are the values for the prototypes for d stimulus features, and v_i are the stimulus dimension on which the rule is applied.

MODEL	PARAMETERS	
ERMI, MI, PFN, RMC	β	
GCM	c, ε, w_i	$\forall\, i \in \{1,2,\ldots,n\}$
PM	$c, \varepsilon, w_i, q_{1,i}, q_{2,i}$	$\forall\, i \in \{1,2,\ldots,n\}$
RULE	v_1, v_2, ε	
CLAUDE-V2	ε	

Bayesian model comparison

After fitting the model parameters to human data using maximum likelihood estimation, we computed the Bayesian information criterion (BIC), which penalizes model fitting performance based on its complexity, for models m for a given participant as follows:

$$\text{BIC}_m = -2 \cdot \max_{\theta_m} \sum_{t=1}^{T} \log p_{\theta_m}(\hat{y}_t \mid x_{1:t}, y_{1:t-1}) + |\theta_m| \log(T) \tag{6}$$

where $|\theta_m|$ is the number of parameters estimated for the model m, T is the number of trials in the task and $\hat{y}_t$ is the choice made by the participant in a given trial t.

Once computed, we compared the goodness-of-fit between models using posterior model frequency, which measures how often a given model offers the best explanation in the population. For computing it, we used a Python implementation of the Variational Bayesian Analysis (VBA) toolbox[21]. The toolbox required providing log-evidences for each model and participant pair, which we approximate using $-0.5 \cdot \text{BIC}_m$; see Rigoux et al. study[22] for details about this model comparison procedure.

Table S3. Mean performance of humans and models for each rule type in replication of[11] study over 15 blocks. Human data was taken from Table 1 in[10].

Model	**Rule**						**MSE**
	Type 1	*Type 2*	*Type 3*	*Type 4*	*Type 5*	*Type 6*	
Humans	.0201	.0565	.1015	.1120	.1212	.2048	.0000
ERMI	.0586	.0891	.0855	.0826	.0888	.1172	**.0287**
MI	.0686	.4089	.2404	.1431	.2880	.4201	.2627
PFN	.0170	.3405	.1533	.0226	.2371	.3975	.1736
RMC	.1329	.2215	.1903	.1718	.2132	.3364	.1003

CLAUDE-V2 as a cognitive model of human category learning

To simulate the study by Badham et al.[12] using CLAUDE-V2, we queried the model with the prompt shown below. Geometric stimuli from the original experiment were described in text format. The order of presentation of the stimulus was randomized and the number of presentations per block was compared to the original study. As the Claude API returns only sampled tokens,

not log-probabilities, we coded predictions as binary outcomes, $\pi(k=1 \mid x_t; x_{1:t-1}, y_{1:t-1})$. The final model predicted category probabilities is again a mixture between the category prediction from the model and a random guess as mentioned in Equation 7. We conducted 96 simulation runs for each of the six categorization rules.

$$p(k=1 \mid x_t) = (1-\varepsilon)\pi(k=1 \mid x_t; x_{1:t-1}, y_{1:t-1}) + \varepsilon \cdot \mathrm{K}^{-1} \tag{7}$$

Prompt for Badham et al. 2017 study

In this experiment, you will be shown examples of geometric objects. Each object has three different features: size, color, and shape. Your job is to learn a rule based on the object features that allows you to tell whether each example belongs in the {A} or {B} category. As you are shown each example, you will be asked to make a category judgment and then you will receive feedback. At first you will have to guess, but you will gain experience as you go along. Try your best to gain mastery of the {A} and {B} categories.

- In trial 1, you picked category {A} for Big Black Square and category {A} was correct.
- In trial 2, you picked category {A} for Small Black Triangle and category {B} was correct

Human: What category would a Small Black Triangle belong to? (Give the answer in the form "Category ⟨your answer⟩").
Assistant: Category

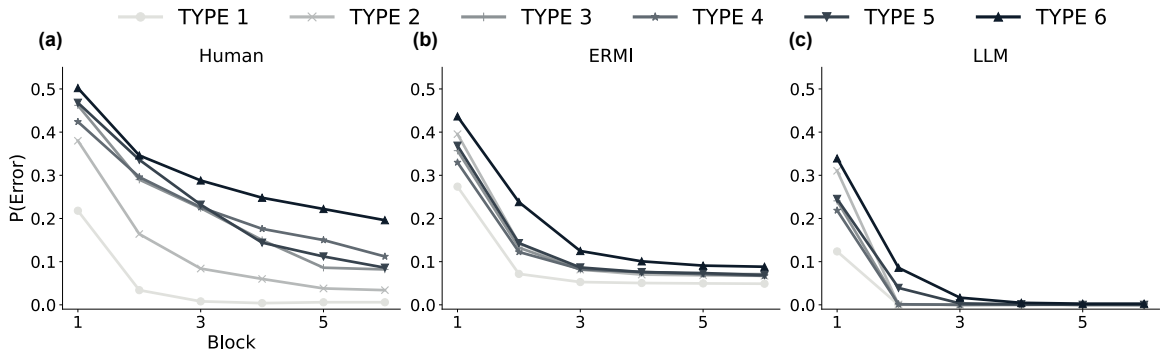

Figure S5. Unlike ERMI, CLAUDE-V2 does not show human-like learning difficulties: **(a-c)** Average error probabilities for each task *type* in each block of 16 trials for (a) humans, (b) ERMI, and (c) LLM. Human data in (a) was reproduced from Table 1 in Nosofsky et al.[10] study. ERMI was simulated on *type 1-6* tasks for 50 runs with the inverse temperature set to $\beta = 0.4$. CLAUDE-V2 was simulated for 94 runs each on *type 1-6* tasks with temperature term set to 0.

Decision Making

LLM-generated tasks

Prompts: In the following, we provide the prompts used in the two stages of decision making learning domain. We used the following prompt to synthesize the names of stimulus features and targets, similar to function learning, separately for each of the three conditions, ranking, direction, and unknown.

Synthesize stimulus feature name and its target for ranking condition

I am a psychologist who wants to run a function learning experiment. In a function learning experiment, a real-world feature is mapped to its corresponding target, with both feature and target taking on continuous values.

Please generate names for features and its corresponding target for 250 different function learning experiments. Additionally, order the feature names according to how well they predict the target:

– feature name, target name

Synthesize stimulus feature name and its target for direction condition

I am a psychologist who wants to run a function learning experiment. In a function learning experiment, a real-world feature is mapped to its corresponding target, with both feature and target taking on continuous values.

Please generate names for features and its corresponding target for 250 different function learning experiments. Additionally, the features should be such that higher feature values lead to higher target values:

– feature name, target name

Synthesize stimulus feature name and its target for unknown condition

I am a psychologist who wants to run a function learning experiment. In a function learning experiment, a real-world feature is mapped to its corresponding target, with both feature and target taking on continuous values.

Please generate names for features and its corresponding target for 250 different function learning experiments:

– feature name, target name

Next, we prompted the LLM to generate values for tasks generated from stage 1:

Generate values for ranking condition

I am a psychologist who wants to run a function learning experiment. For a function learning experiment, I need a list of features with their corresponding target. The features in this case are feature1, feature2, feature3, and feature4. These features take on only numerical values and must be continuous. The target, <target>, should be predictable from the feature values and must also have continuous values. Note that the features are listed according to how well each of them can predict the target. The first feature is most useful for predicting the target, the second feature is the second most useful, etc.
Please generate a list of <num-data> feature-target pairs sequentially using the following template for each row: - feature value 1, feature value 2, feature value 3, feature value 4, target value

Generate values for direction condition

I am a psychologist who wants to run a function learning experiment. For a function learning experiment, I need a list of features with their corresponding target. The features in this case are feature1, feature2, feature3, and feature4. These features take on only numerical values and must be continuous. The target, <target>, should be predictable from the feature values and must also have continuous values. Note that the values taken by the features should be such that higher feature values lead to higher target values.
Please generate a list of <num-data> feature-target pairs sequentially using the following template for each row: - feature value 1, feature value 2, feature value 3, feature value 4, target value

Generate values for unknown condition

I am a psychologist who wants to run a function learning experiment. For a function learning experiment, I need a list of features with their corresponding target. The features in this case are feature1, feature2, feature3, and feature4. These features take on only numerical values and must be continuous. The target, <target>, should be predictable from the feature values and must also have continuous values.
Please generate a list of <num-data> feature-target pairs sequentially using the following template for each row: - feature value 1, feature value 2, feature value 3, feature value 4, target value

Parsing and pre-processing: The parsing expressions used and the data preprocessing steps are the same as in the function learning domain.

Qualitative analysis of synthesized input features and labels: We show the counts for the top-50 most frequently occurring names for (a) input features and (b) targets in Figure S6. We found that the LLM tends to produce input-target pairs that are relevant to everyday life such as supply-demand influence on productivity, diet-genetics influence on weight change, cloud cover-humidity on crop yield, study time-intelligence quotient on test score, etc.

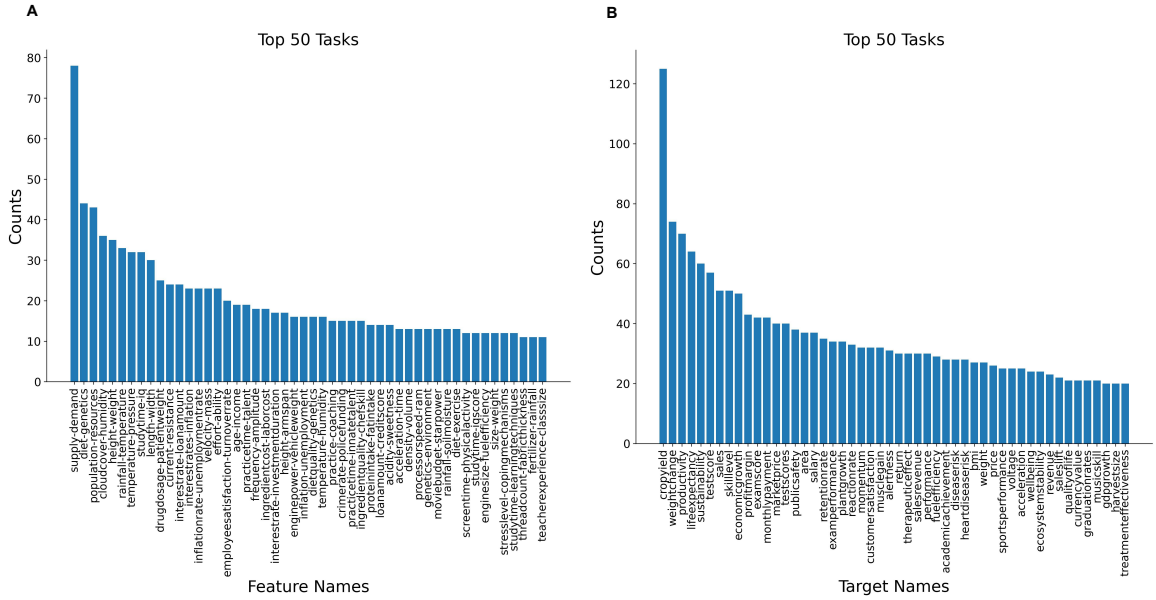

Figure S6. Frequency of input and target labels in CLAUDE-V2 synthesized decision making tasks: Counts for the top-50 most frequently occurring (a) two-dimensional input feature names and (b) target names computed over 9254 LLM-generated decision learning tasks belonging to the unknown condition. These distributions confirm that the LLM-generates real-world functional relationships that are useful for everyday decision making.

Data-distributional properties of LLM-generated tasks: We generate three datasets of decision-making tasks, one for each of (A) unknown, (B) ranking, and (C) direction, following the prompts described above. To examine their properties and verify if the manipulation was successful, we computed four key statistics: input correlations, sparsity in predictive features, ranking of feature importance, and directionality of features with respect to the target, and compared them across datasets. Specifically, we contrasted the ranking and direction conditions with the unknown condition, which served as a baseline. We found that the first feature was more often the most important feature in terms of predictive power (see the caption of Figure S7 for details on the calculation) in the ranking condition (51.76%) than in the unknown condition (43.75%). Likewise, the proportion of features positively correlated with the target was higher in the direction condition (92.46%) than in the unknown condition (79%).

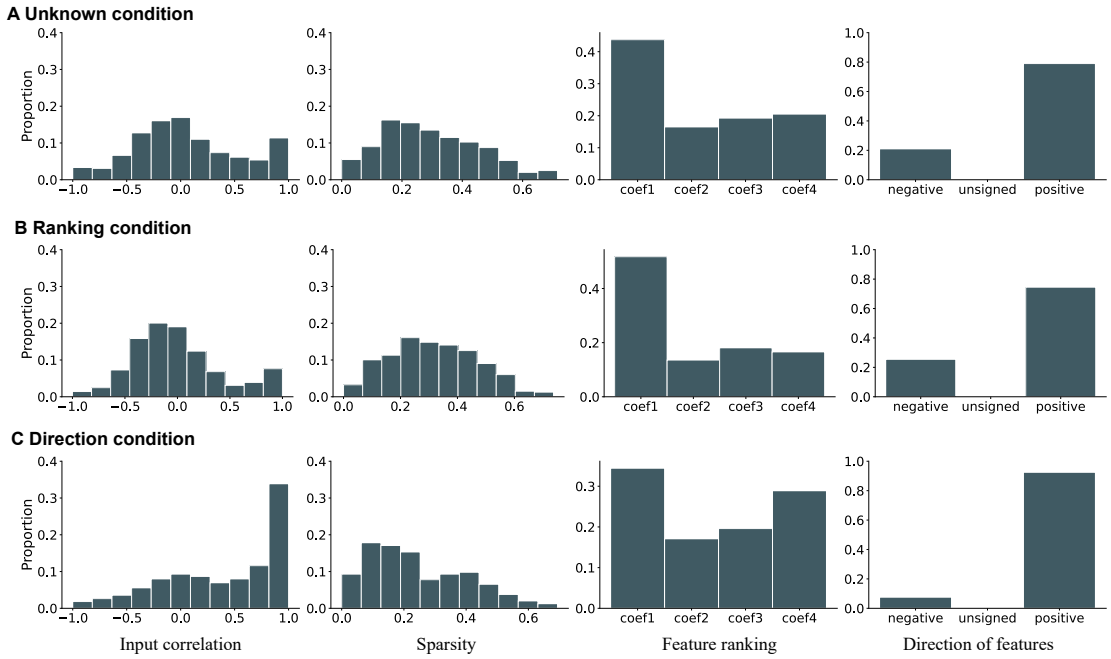

Figure S7. Data-distributional properties of LLM-generated decision making tasks for (A) unknown, (B) ranking and (C) direction condition. Histogram of Pearson correlation coefficients between all distinct pairs of normalized input features (first column). Histogram of Gini coefficients computed on the absolute ordinary least squares (OLS) weights when regressing the normalized target on all normalized inputs with an intercept (second column; higher values indicate sparser weights). Histogram of the index of the input feature with the largest absolute per-feature OLS weight, where each per-feature model regresses the target on a single feature with an intercept (third column; feature ranking). Histogram of the sign of per-feature OLS weights from those single-feature-with-intercept regressions (fourth column; direction).

Human studies

Binz et al. 2022[23]. This study was conducted on 27 participants in total, with each participant performing 30 different paired comparison tasks. Tasks were generated by first sampling feature weights from a standard normal distribution. Feature vectors for each option were then drawn from a multivariate normal distribution with zero mean and fixed covariance. Finally, the binary choice outcome was determined by sampling from a Bernoulli distribution, where the success probability was given by a probit regression over the difference in feature values (see Equation 2 in the main paper). The feature weights were kept the same within a task, which consisted of 10 trials, but were resampled between tasks. All participants performed the same set of paired comparison tasks but presented in randomized order. In Experiment 3a, participants observed two features per option, whereas in Experiment 3b they observed four features per option. In neither of these two experiments, information about the ranking of the features and their directions were provided. The experiment itself was framed as an alien sports competition on an unknown planet. Participants observed two or four numerical attributes for two aliens, depending on the experiment

they were part of. They indicated their choice by pressing a button corresponding to the alien they believed would most likely win. This cover story was used so that the meaning of the feature attributes remained abstract for each participant. Participants were not told about the underlying feature weights, and they had to learn them through trial and error, using the feedback about correct choice provided after each trial. All participants in the experiment performed a short tutorial and went through a comprehension check, which ensured clear understanding of the experimental protocol before data-collection.

Hand-crafted tasks

Synthetic paired-comparison tasks used for training MI: We generated three synthetic datasets of paired-comparison problems (between 7000-9000 tasks per set) under *ranking*, *direction* and *unknown* conditions. For each task, a weight vector $w \in \mathbb{R}^d$ was sampled from a standard normal distribution. In the *direction* condition, weights were constrained to be non-negative by taking absolute values; in the *ranking* condition, feature importance was rank-ordered by sorting weights by magnitude; and in the *unknown* condition, weights were left unconstrained. To generate options, the feature vectors were sampled from a zero-mean multivariate normal distribution with covariance $\Sigma = L\,\mathrm{diag}(\theta)\,L^\top$, where L was drawn from an LKJ (Lewandowski–Kurowicka–Joe distribution; $\eta = 2$) prior and $\theta = \mathbf{1}$. The LKJ distribution is a flexible prior over correlation matrices that allows control over the strength of correlations while ensuring positive definiteness. Each trial presented a pair of options $x_a, x_b \sim \mathcal{N}(0,\Sigma)$, with the comparison input defined as $x = x_a - x_b$. We randomly determine which option has the highest criterion by sampling from a Bernoulli distribution as follows: $y \sim \mathrm{Bernoulli}\big(\Phi(w^\top x/\sigma)\big)$ with $\sigma = 0.1$. Each task contained a maximum of 10 trials, which corresponded to the length of the experiment in which this model was evaluated.

Model architecture, and training

The input vector for a given trial in a decision making task was the difference between the input features for the two options, computed for each dimension independently, and the correct target option from the previous trial. The number of features in the decision making task was either two or four dimensions and the total number of observations in a given task was 20. These inputs were embedded into a 64-dimensional space, with positional encoding applied using sine and cosine functions of varying frequencies, following Vaswani et al.[6]. A causal attention mask ensured that predictions at each time step were conditioned only on all previous inputs. These masked sequences were processed using a Transformer decoder composed of six layers, with 64-dimensional embeddings, eight attention heads, and 256 hidden units in the feedforward layers. The decoder outputs were passed through a linear projection to produce weights for the different feature dimensions. The likelihood of a target option is then calculated by first projecting the output through a linear layer, multiplying it element-wise with the current input features, summing across dimensions, and passing the result through a sigmoid to obtain a Bernoulli probability. Training was performed using the negative log-likelihood (NLL) loss over all input observations in a batch. The model parameters were updated as mentioned before using the ScheduleFree optimizer[7] with a baseline learning rate of 3×10^{-4}. Each model was trained for 100 000 episodes, with periodic evaluation on held-out tasks to monitor generalization performance.

Baseline models

Apart from the MI model derived by meta-learning on tasks generated with hand-crafted priors, we considered three other cognitive models as baselines in the domain of decision making, as detailed below.

Single-cue decision maker (SC): In Equation 8, we demonstrate formally how the heuristic of single-cue decision making makes a decision given the input feature. Note that x^* indicates that the model only takes into account a single feature, which in this case was the most predictive feature. This means that only one parameter is fitted to human choices.

$$p(y_t = 1 \mid x_t, \theta_m, m = \mathrm{SC}) = \Phi\left(\frac{\theta_m \cdot x_t^*}{\sqrt{2}\sigma}\right) \tag{8}$$

where Φ is the cumulative distribution function of a standard normal distribution, θ_m is the weight of the selected feature, and σ is the noise standard deviation.

Equal weighting decision maker (EW): We considered a probabilistic version of the equal weighting model, as shown in Equation 9. When $w > 0$, this model probabilistically selects the option with the larger sum of features. In contrast, when $w < 0$, it selects the option with the smaller sum of features. Once again, only one parameter is fitted to the human data.

$$p(y_t = 1 \mid x_t, \theta_m, m = \mathrm{SC}) = \Phi\left(\frac{\theta_m \cdot \sum_{i=1}^{d} x_{t,i}}{\sqrt{2}\sigma}\right) \tag{9}$$

where Φ is the cumulative distribution function of a standard normal distribution, θ_m is the feature weight, and σ is the noise standard deviation.

Feedforward neural network (NN): We used a feedforward neural network from the Binz et al.[23] study as an additional baseline model. This model predicts the target given the difference between the input features of the two options and the previous target as input. The network consisted of a single hidden layer with 128 units followed by two linear transformations projected to the mean and (log) standard deviation of a normal distribution. The neural network parameters were trained by gradient descent on the negative log-likelihoods of the target. During model fitting, the learning rate parameter and the inverse temperature term were fit to human choices; see Appendix F in Binz et al.[23] for implementation details.

Model fitting and comparison

For fitting the model parameters, we performed the maximum likelihood estimation using Bayesian optimization[24], following the procedure used by Binz and colleagues.[23]. A complete list of model parameters that are fitted to human choices can be found in Table S4. Upon fitting, we followed the same exact steps as described above for category learning for Bayesian model comparison. That is, we used a VBA tool box, where we provide $-0.5 \cdot \mathrm{BIC}_m$ as an approximation of log-evidence for each model and participant; see Rigoux et al. study[22] for details.

Table S4. This table provides the complete list of model parameters that were fit to human data in the domain of category learning, where β is the inverse temperature term, θ indicates the weights for the stimulus feature dimension, and α is learning rate.

MODEL	PARAMETERS
ERMI, MI	β
SC	θ
EQ	θ
NN	α, β

Alternative LLMs

During the early stages of this work, we also considered two other LLMs: Llama-2[25] and GPT-4[26], which were among the best performing models at the time. However, the non-instruction-tuned Llama-2 (the only version available at the time) could not consistently produce the 100+ data points required for each category learning task. Its outputs were also difficult to parse, as they frequently failed to follow the specified format. More recently, with Llama-3.1 (70B)[27], we were able to generate decision-making datasets whose quality matched those produced by CLAUDE-V2.

Preliminary analysis with GPT-4 revealed that it often sampled input features from a uniform distribution, relying on its internal coding module. It also tended to generate only simple heuristic rules, such as requiring the sum of two features to exceed the third, or the mean of two features to be greater than another, for assigning an input to its category. Furthermore, statistical analysis on a small GPT-4 generated dataset showed that its task statistics closely resembled those of category learning tasks with hand-crafted priors (specifically Bayesian logistic regression prior). Due to this lack of diversity in the generated task statistics, we decided to use CLAUDE-V2 over GPT-4.

References

1. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17**, 261–272, DOI: 10.1038/s41592-019-0686-2 (2020).
2. Kwantes, P. J. & Neal, A. Why people underestimate y when extrapolating in linear functions. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 1019 (2006).
3. Little, D. R., Shiffrin, R. M. & Laham, S. M. Function estimation: Quantifying individual differences of hand-drawn functions. *Mem. & Cogn.* **Springer**, 1–20 (2024).
4. Tummers, B. Datathief iii http://datathief. org. *Datathief is a program used to reverse engineer data points from a graph* (2006).
5. Lucas, C. G., Griffiths, T. L., Williams, J. J. & Kalish, M. L. A rational model of function learning. *Psychon. bulletin & review* **22**, 1193–1215 (2015).
6. Vaswani, A. *et al.* Attention is all you need. *Adv. neural information processing systems* **30** (2017).
7. Defazio, A. *et al.* The road less scheduled. *Adv. Neural Inf. Process. Syst.* **37**, 9974–10007 (2024).
8. Devraj, A., Zhang, Q. & Griffiths, T. The dynamics of exemplar and prototype representations depend on environmental statistics. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 43 (2021).
9. Johansen, M. K. & Palmeri, T. J. Are there representational shifts during category learning? *Cogn. Psychol.* **45**, 482–553, DOI: 10.1016/s0010-0285(02)00505-4 (2002).
10. Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C. & Glauthier, P. Comparing models of rule-based classification learning: a replication and extension of shepard, hovland, and jenkins (1961). *Mem. Cogn.* **22**, 352–369, DOI: 10.3758/bf03200862 (1994).
11. Shepard, R. N., Hovland, C. I. & Jenkins, H. M. Learning and memorization of classifications. *Psychol. Monogr. Gen. Appl.* **75**, 1–42, DOI: 10.1037/h0093825 (1961).
12. Badham, S. P., Sanborn, A. N. & Maylor, E. A. Deficits in category learning in older adults: Rule-based versus clustering accounts. *Psychol. Aging* **32**, 473–488, DOI: 10.1037/pag0000183 (2017).
13. Smith, J. D. & Minda, J. P. Prototypes in the mist: The early epochs of category learning. *J. Exp. Psychol. Learn. memory, cognition* **24**, 1411 (1998).
14. Medin, D. L. & Schaffer, M. M. Context theory of classification learning. *Psychol. Rev.* **85**, 207–238, DOI: 10.1037/0033-295x.85.3.207 (1978).
15. Nosofsky, R. M., Palmeri, T. J. & McKinley, S. C. Rule-plus-exception model of classification learning. *Psychol. Rev.* **101**, 53–79, DOI: 10.1037/0033-295x.101.1.53 (1994).
16. Müller, S., Hollmann, N., Arango, S. P., Grabocka, J. & Hutter, F. Transformers can do bayesian inference. *arXiv preprint arXiv:2112.10510* **arXiv** (2021).
17. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, 8024–8035 (Curran Associates, Inc., 2019).
18. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
19. Anderson, J. R. The adaptive nature of human categorization. *Psychol. Rev.* **98**, 409–429, DOI: 10.1037/0033-295X.98.3.409 (1991).
20. Nosofsky, R. M. Attention, similarity, and the identification–categorization relationship. *J. experimental psychology: Gen.* **115**, 39 (1986).
21. Daunizeau, J., Adam, V. & Rigoux, L. Vba: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS computational biology* **10**, e1003441 (2014).
22. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies—revisited. *Neuroimage* **84**, 971–985 (2014).
23. Binz, M., Gershman, S. J., Schulz, E. & Endres, D. Heuristics from bounded meta-learned inference. *Psychol. review* **Psychological review** (2022a).
24. Team, T. G. Gpyopt: A bayesian optimization framework in python. http://github.com/SheffieldML/GPyOpt (2016).
25. Touvron, H. *et al.* Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

26. Achiam, J. *et al.* Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

27. Meta Platforms, I. Llama 3.1 70b model (2024).

What did we learn?

Outlook 3

3.1 Discussion

The ability of humans to learn and adapt is extraordinary. From acquiring proficiency in a foreign language to gaining mastery of your professional field, learning is central to many key aspects of everyday life. It would be close to impossible to thrive in an ever-changing world without this propensity. A wide range of frameworks have been proposed to understand the factors that drive human learning: from rational analysis to resource-rational analysis and ecological rationality. However, modeling approaches used to test theories under these frameworks have faced issues related to their scalability, flexibility, reliance on hand-crafting, and neural plausibility, preventing them from providing a holistic account of human learning. In this thesis, we proposed that meta-learning offers a radical new approach to building computational models of human learning that addresses these shortcomings. To demonstrate it, I first showed how the framework of meta-learning allows for the construction of rational, resource-rational, and ecologically rational learning algorithms, without requiring careful handcrafting. Also, I looked at how it combines the strengths of Bayesian and connectionist models while remaining scalable, flexible, bounded, and adaptive. We then used it to perform rational analysis of optimism bias, resource-rational analysis of zero-shot compositional reinforcement learning, and ecological rational analysis of human function learning, category learning, and decision-making. Across these different studies, I found that meta-learning offers a domain-general framework to test theories of human learning, unifying different frameworks and modeling approaches.

Distilling symbolic priors into sub-symbolic systems

Def.: Inductive bias

Inductive bias refers to assumptions that guide a learner toward sample-efficient learning and generalization [57]

Symbolic systems, such as Bayesian models, have been particularly successful in capturing rapid learning and generalization in humans [61, 137]. A key reason for their success is that Bayesian models effectively encode human inductive biases, typically specified by defining a probability distribution over hypothesized concepts [57]. In contrast, sub-symbolic systems like neural networks require significantly more training examples due to their relatively weak inductive biases. While they excel at a wide range of tasks, including real-world image classification [138], their generalization performance worsens substantially when trained on small datasets [139].

Meta-learning helps bridge the gap between symbolic and sub-symbolic systems. By enabling the distillation of symbolic priors from Bayesian models into neural networks, it allows neural architectures to achieve sample-efficient learning comparable to symbolic approaches [140]. For example, in Section 2.4 – Human-like category learning by injecting ecological priors from large language models into neural networks, I meta-learned on samples drawn from the rational model of categorization

[6] and showed that the resulting model not only replicated the learning difficulty curves exhibited by participants (and the original model) but also captured human behavior well. Later, I showed that this approach extends to other domains, such as function learning. In Section 2.5 – Meta-learning ecological priors from large language models captures human learning and decision making, I meta-learned on functions sampled from the prior used by the rational model of function learning [61], and found that the resulting model captured human behavior well. Finally, I demonstrated how the approach also supports more abstract priors, such as the propensity to reason compositionally over previously learned reward functions; see Section 2.3 – A resource-rational account of zero-shot compositional inference in a reinforcement learning setting for details.

The meta-learned rational model of categorization was the second-best model in terms of matching human learning curves in simulations, and in model comparison on the human dataset from Badham et al. 2017 [136]. Similarly, the meta-learned rational model of function learning ranked second in fitting both qualitative human behavior and model comparison on the dataset from Little et al. 2024 [141]. In both domains, the best-performing model was the ecologically rational meta-learned inference; see Section 2.5 for full results.

From a practical standpoint, this approach offers several benefits: (1) it leverages the fact that sampling from a generative model is often easier than performing inference; (2) it reduces the need for hand-designed architectures, allowing off-the-shelf neural networks to incorporate symbolic priors; (3) it enables faster inference, once trained, the model runs in a fraction of the time required by traditional Bayesian models. Crucially, it also allows discrete symbolic hypotheses to be expressed as continuous vector-based representations, providing the means to study how high-level inductive biases might be represented in the brain; see Section 3.2 – Limitations and future directions for a detailed discussion.

Nevertheless, it is important to emphasize that distilling symbolic priors into neural networks does not answer a key question: Where do priors come from? I have addressed this question in several of my publications[1], and I will elaborate on it in the following sections.

1: See Section 2.4, Section 2.5, and Section 2.3

Modeling human compositional reasoning using neural networks

The ability to learn new concepts and combine them in flexible ways to solve novel tasks lies at the core of human cognition [139, 142]. For example, if a person understands the meaning of the sentence "John loves the girl," one can reasonably expect them to infer the meaning of "The girl loves John." However, classical connectionist models have historically struggled to perform such systematic generalization.

Building on this observation, Fodor and Pylyshyn argued that connectionist models should not be considered valid cognitive models. Their reasoning was as follows: since neural networks learn distributed representations, they lack a clear internal structure necessary for systematic generalization. The encodings they learn contain overlapping information and are difficult to parse unambiguously into constituent parts. As a result, these models struggle to reuse decomposed components across contexts or recombine them in novel ways to construct richer representations. This led Fodor and Pylyshyn to reject connectionism as a viable framework for studying cognition [143].

[143]: Fodor et al. (1988), 'Connectionism and cognitive architecture: A critical analysis'

In the decades that followed, researchers made significant efforts to address this sharp criticism leveled at connectionist models. While some explored complex network architectures capable of supporting systematic compositionality [144–146], others questioned the premise

itself by arguing that human compositional reasoning is flawed and that people often fail to exhibit systematic generalization [147, 148].

In Section 2.3 – A resource-rational account of zero-shot compositional inference in a reinforcement learning setting in Publications, I have shown that sophisticated architectures are not necessary to get neural networks to reason compositionally. Instead, what is required is a training procedure that explicitly encourages compositional reasoning. Specifically, I demonstrated that a carefully designed learning curriculum that incentivizes compositional reasoning during meta-learning is sufficient to enable an off-the-shelf neural network to generalize systematically. The curriculum required the network, here long short-term memory (LSTM) network, to combine the reward functions learned from the first two subtasks to solve the third compositional subtask. Under this setup, the LSTM performed near optimal zero-shot compositional inference, matching the performance of the compositional Gaussian process regression model– the true generative model underlying the task. In parallel, Lake and Baroni introduced an approach, called meta-learning for compositionality (MLC), that similarly encourages systematic generalization. They showed that MLC can also guide a standard neural network (there a transformer model [44]) to perform a human-like systematic generalization in language-based reasoning tasks [149].

Furthermore, I extend the meta-learning for compositionality approach to account for deviations from optimal compositional inference observed in humans in Section 2.3. I attributed the departure from optimality to resource constraints, such as working memory and attentional limitations, with which humans contend [7]. To model this, I modified the meta-learning procedure so that it enforces a constraint on the (algorithmic) complexity of the emerging in-context learning algorithm [53]. Specifically, I limited the description length of the neural network, which is the number of bits required to represent its weights, by including the Kullback-Leibler divergence term that penalizes the weights from deviating from unit normal into the meta-learning objective; see Equation 17 in Section 2.3. By controlling the description length, I derived resource-rational learning algorithms that perform compositional reasoning to varying degrees. I found that such constrained learning algorithms capture the behavior of the participants better than seven other cognitive models, including symbolic models such as Gaussian process regression and Bayesian mean tracker, in two different experiments. Additionally, the description length parameter reliably quantified individual differences in participants' capacity for zero-shot compositional reasoning.

Together, these findings suggest that meta-learning, when combined with learning curricula with the right incentive structure and cognitively plausible constraints, offers a promising path toward resolving the systematicity debate in neural networks and demonstrates that connectionist models remain a viable framework for studying cognition.

Rational analysis of biological and artificial systems

The strength of rational analysis lies in its ability to explain "why cognition works, by viewing it as an approximation to ideal statistical inference given the structure of natural tasks and environments" [150]. When

applied to the behavior of a system—whether natural or artificial—it offers a principled lens to uncover the underlying factors that drive behavior, be it environmental structure, cognitive limitations, or task objectives.

In Section 2.2 – In-context learning agents are asymmetric belief updaters, I used meta-learning for a rational analysis of the optimism bias displayed by humans in a reinforcement learning setting. Until our study, prior work had primarily relied on simulation-based analyses within restricted model classes, as deriving a fully rational solution by hand was intractable for this task [125, 151]. To address this limitation, I used meta-learning to derive rational agents for bandit tasks used by previous studies to study human reinforcement learning [124, 125]. I did this by meta-learning an agent on distribution of bandit tasks, where reward probabilities for the arms of the bandit tasks were sampled independently for each arm from a uniform distribution at the start of each episode. This setup captures the assumption that individuals begin each task uncertain about reward probabilities. After training, we found that the resulting model implemented a Bayes-optimal policy on the task configuration given to human participants. When we analyzed the simulated behavior of this model using Rescorla-Wagner model, akin to standard practice in optimism bias studies , we observed that the optimal agent displayed asymmetric belief updating—mirroring human behavior—with learning rates for positive prediction errors significantly larger than those for negative ones. This finding suggests that optimism bias emerges as a rational strategy, shaped by optimal adaptation to the diverse environments previously observed or interacted with.

In Section 2.2, I also analyzed the behavior of large language models, shown to be capable in-context reinforcement learners [152], through the lens of rational analysis. To this end, I adapted the human experiment for LLMs by converting the task instructions into natural language and querying the LLM trial-by-trial to choose between the arms of the bandit, given the history of its previous choices and their outcomes. Consistent with previous findings in humans and rational models derived using meta-learning, the LLM learned more from positive than negative prediction errors. However, the asymmetry in LLMs was much greater than those observed in humans and meta-learned models, suggesting a similar but more pronounced bias toward learning from positive prediction errors.

These findings demonstrate the utility of using meta-learning to train idealized learning algorithms, particularly in settings where deriving fully rational models is intractable, thereby enabling a rational analysis of behavior, whether natural or artificial.

A principled approach to derive decision-making heuristics

Ecological rationality lends support to the idea that heuristics, which are simple resource-rational strategies adapted to the properties of the environment, explain how people make decisions in everyday life. However, the current approach to deriving heuristics is quite laborious and unsystematic; see Section 1.7 in Background for detailed reasons.

In Section 2.5 – Meta-learning ecological priors from large language models captures human learning and decision making, we proposed a new framework, called ecologically rational analysis, which provides a principled approach to derive heuristics that are rationally adapted to the statistics of the real world environment. The proposed framework achieved this by making use of two recent developments in machine learning: (1) large language models [153] (2) meta-learning [154].

First, LLMs are trained on Internet scale human – and partly machine – generated text. This has allowed them to successfully generate realistic tabular datasets [155] and internalize everyday task distributions similar to humans [156]. We showed that LLMs not only capture traces of human thought and reasoning [157] but also capture the statistics of the world they inhabit. Specifically, we leveraged LLM to generate datasets of cognitive tasks that capture ecologically valid statistics. As a result, we can now model properties of the real world across different cognitive domains at high throughput and with minimal human intervention, as long as the output can be expressed in natural language. Second, meta-learning offers a principled approach to deriving learning algorithms adapted to the statistics of the tasks sampled from a prespecified training distribution. Assuming that the training distribution captures statistics of the real-world environments, the resulting learning strategies are adapted to the ecological priors captured by it. Therefore, meta-learning provides a useful framework for investigating the emergence of different heuristics and how they could be learned through experience.

It should be noted that the current approach to meta-learning excludes other contributing factors, such as evolution, culture, and social learning, driving the acquisition and application of heuristics[110] and must be considered in future work [158].

We then successfully evaluated this approach in practice. Prior work has shown that people adopt different decision-making heuristics depending on the statistical properties of the task [159]. For example, when participants were informed about the importance of attributes but not the direction of their correlation with the target (ranking condition), they used a one-reason decision strategy. When the direction was known but not the ranking (direction condition), they relied on an equal-weighting strategy. When neither ranking nor direction was known (unknown condition), they used a weighted combination of attributes to guide their choices (see Section 2.5 for additional details). We showed that Claude-v1 can be prompted to generate decision-making problems corresponding to these three conditions. Unlike synthetically generated tasks, LLM-generated problems capture additional statistical properties, such as the extent of correlation, sparsity, and value distributions, without requiring manual tuning. We then trained models via meta-learning on each task distribution and found the learned strategies closely matched those used by humans: a single-cue strategy in the ranking condition, an equal-weighted strategy in the direction condition, and a weighted combination in the unknown condition.

These results suggest that ecologically rational analysis offers the means to answer questions such as which heuristic to apply depending on the properties of the environment and how to dynamically adapt its strategy when these properties change and based on the participants' history of previous interactions. Even though its approach differs from those used by classical ecological rationality, its goal remains the same: to find an adaptive rationale behind the strategies used by people in any given environment.

Ecological priors captures a substantial amount of variance in human learning and decision-making

In Section 2.5 – Meta-learning ecological priors from large language models captures human learning and decision making, we introduced a new computational framework, called ecologically rational analysis, to answer the question of how much variance do ecological priors capture in human learning and decision-making. By using large language models to automatically generate ecologically valid task environments, and meta-learning to distill priors from task statistics, it allowed for an automated derivation of computational models that approximate the ideal statistical inference under the structure of natural tasks and environments [150]. The resulting class of models, called Ecological Rational Meta-learned Inference (ERMI), therefore, provided the means to examine the mechanism and extent to which environmental properties drive human learning and decision-making.

Across 15 experiments that span three cognitive domains, namely function learning, category learning, and decision-making, we found that ERMI accounts for a substantial amount of variance in human behavior. Not only did it capture classic behavioral signatures in each domain, but it also made precise prediction of trial-level human choices compared to many established cognitive models.

First, ERMI reproduced many characteristic behaviors displayed by humans during function learning. Namely, ERMI (1) found positive linear functions easier than negative linear functions [160–162]; (2) learned these functions faster [163, 164]; (3) acquired monotonic functions better than nonmonotonic functions [160, 164, 165]; (4) found cyclic functions more challenging than noncyclic functions [164]; (5) generalized better during interpolation than during extrapolation [163, 165, 166]; (6) and showed a bias toward zero offset when generalizing linear positive functions [167]. Additionally, ERMI also fit the functions people drew to capture the true causal function underlying scatter plots shown to them in both the interpolation and extrapolation conditions better than the rational model of function learning.

Second, during category learning, ERMI captures three patterns that have been observed in humans: (1) it found the same category structures difficult that human found difficult, replicating the canonical findings from Shepard and colleagues [14], (2) it showed the same shift in learning strategy as people, changing from prototype-based to exemplar-based strategy with experience [133]; (3) it displayed generalization patterns similar to people when queried on unseen stimuli [168]. Furthermore, ERMI explained trial-level human choices better than eight established cognitive models in two different human experiments [135, 136].

Third, ERMI flexibly shifts between heuristics, for example, from single-cue decision-making to an equal-weighting strategy, depending on the statistics of the underlying task, just as humans do. It also captured human decisions at the trial-by-trial level in two human experiments, where the underlying statistics was unknown and the dimensions of the stimulus features varied between two and four [22].

Together, these findings demonstrate that ERMI–a computational model derived by extracting ecological priors from large language models and

distilling them into neural networks using meta-learning–sufficiently accounts for much of human learning and decision-making.

Towards bounded ecologically rational analysis

Throughout this thesis, I have illustrated how meta-learning offers an alternative approach to building computational models that align with theoretical frameworks, such as rational analysis, resource-rational analysis, and ecologically rationality. I have also discussed its advantages over conventional approaches and the unique benefits it offers, from the ease of manipulating computational resources to integrating neuroscientific insights into network architecture. However, I argue that the key strength of meta-learning lies not merely in enabling the testing of theories of human learning and decision-making within these frameworks, but in providing a unifying foundation that bridges them.

More concretely, I have investigated the respective contributions of ecological adaptation and bounded resources to human learning using meta-learning (see Section 2.4, Section 2.5, and Section 2.3). By imposing constraints on the computational and algorithmic complexity of the network, and using LLMs to control the data-distributional properties of the training environment, I have shown that it is possible to derive computational models that are optimally adapted either to limited resources or to ecologically valid task statistics. But what if we manipulated both of these factors jointly rather than independently? This would open up new possibilities for studying, in silico, how learning strategies emerge as a joint adaptation to cognitive limitations and ecological environments.

To achieve this, I propose a new framework called bounded ecologically rational analysis[2] , which unifies the advantages of rational analysis, resource-rational analysis, and ecological rational analysis. This framework aims to automate the derivation of computational models that implement resource-rational strategies directly adapted to the statistical structure of natural tasks and environments. The resulting class of models, called Bounded Ecological Rational Meta-learned Inference (BERMI), provide the means to examine how resource constraints and ecological adaption jointly shape human cognition.

2: To the best of my knowledge, Lewis, Howes, and Singh's Ecological-bounded-optimality framework comes the closes to bounded ecologically rational analysis. Citing its potential, the authors say that "Ecological-bounded-optimality explanations offer perhaps the richest classes of psychological theory, although there are fewer clear examples in the literature in this type" [92].

In an ongoing work, I am investigating the potential of models constructed using bounded ecologically rational analysis to explain human category learning, with promising early results. Specifically, I derived BERMI using a similar procedure as ERMI, while bounding their computational capacity. I found that BERMI quantitatively explains human choices better than nine established cognitive models, including ERMI, in two different category learning experiments. In addition, it captures several qualitative aspects of human categorization, such as learning difficulty and learning speed, much better than other competing models.

Despite their promise, these results merely scratch the surface of what is possible with a bounded ecologically rational analysis. The generalization of BERMI to other cognitive domains is yet to be evaluated, the contribution of computational resources over data-distributional properties still needs to be analyzed, and finally, the trade-off between the two factors must be investigated more thoroughly. I suspect that the domain of decision-making would be of specific interest to explore these

questions, since resource-efficient and ecologically adapted strategies, like heuristics, are commonplace in this domain. For instance, it would be interesting to examine carefully if and how BERMI can be used to isolate the contributions of ecological adaptation and resource constraints towards learning a heuristic. More broadly, understanding the dynamics of this trade-off and the conditions under which one factor dominates the other remains a key open challenge.

Overall, early results suggest that bounded ecologically rational analysis has the potential to be an influential framework for automated testing of human learning theories. Future work must build on these findings to make this framework a valid in-silico approach to examine the role of limited resources and ecological environments in shaping cognition.

3.2 Limitations and future directions

Despite their success at capturing human learning across multiple domains, meta-learning is yet to deliver on all its promises. Next, I will discuss the current limitations faced by the framework and outline directions for future work to address them. For clarity and readability, this section is organized into four subsections: (1) Richer environments and tasks; (2) More plausible architectures, objectives, and constraints; (3) bridging different levels of analysis; (4) Extending to other animal species.

Richer tasks and environments

The meta-learned models developed across the studies in this thesis were designed to to explain behavior in controlled task setups typical of psychology experiments. These tasks often involve synthetic stimuli, limited action spaces, finite state spaces with known transitions, and relatively simple stimulus-response mappings [22, 53, 169–171]. As a result, their ecological validity [172] and generalizability [173] have recently been subjected to scrutiny. Moreover, meta-learning distributions defined over such simplified environments have limited expressiveness. They are usually confined to tasks belonging to a single task domain that have a predetermined input and target configuration. This leads to the issue that whenever the task domain or input-target configuration changes, the meta-learning distribution must be reconstructed, and the models retrained from scratch. Therefore, the complexity of the model architecture, the representations they can learn, and the behavior they can exhibit have been rather limited.

A simple solution to the limitations discussed above is to meta-learn in a richer representational space—such as language. Language offers incredible flexibility in terms of the diversity of task structures one can specify and the range of outputs one can generate [48]. For example, large language models trained purely on textual corpora have been shown to function as implicit meta-learners capable of performing Bayesian inference in natural language [174]. LLMs have been shown to exhibit rich behavior across diverse task settings, even explaining human behavior quite well off the shelf [175]. Yet, meta-learning only in the language

space does not necessarily solve the problems highlighted above, as they would still lack crucial components such as embodiment, multimodality, and social learning [176].

Meta-learning in naturalistic environments, such as games [177], presents a promising step towards tackling this issue. Recent work has shown that meta-learning is scalable to embodied 3-D worlds [178], making it an appropriate time to pursue this direction. Among the different games available, those generated by procedural means would be of particular value. Procedurally generated games, such as Minecraft [179], not only allow complex behavior within an ecologically valid setting, but are also composable and implementable with computational methods. I hypothesize that meta-learning on diverse tasks within a rich environment allows the acquisition of inductive biases necessary for sample efficient learning and generalization. Assuming that it does, it offers the means to study the factors that facilitate the learning of complex inductive biases, the processes through which it operates, and its mechanistic underpinnings.

More plausible architectures, objectives, and constraints

The neural network architecture at the core of meta-learning places significant constraints on the emerging learning algorithm. For instance, the base architecture in most of my studies has been the transformer model [44]. Although transformers have several advantages in terms of scalability and training stability [48], they have a key limitation: the size of their context window is fixed after training. This places a strict upper bound on the number of observations that the model can process at inference. Recurrent architectures, such as long short-term memory networks, do not suffer from this bottleneck, but are notoriously difficult to train on long sequences due to vanishing and exploding gradient problems. However, recent recurrent architectures, such as xLSTM [180] and state-space models [45], have demonstrated ease of scaling, offering a valid solution to this problem[3]. Future work needs to explore how well meta-learning with these newer architectures can capture key aspects of human learning.

3: Alternatively, Irie, Csordás, and Schmidhuber has recently introduced a modern version of self-referential weight matrices [182], which can be applied recursively without any memory or state overhead to overcome the hard limit imposed on context size in transformer architecture [181].

The choice of network architecture is crucial for yet another reason: the inductive biases of the base network significantly influence how the model learns. In ongoing work, implementing bounded meta-learned inference with a transformer core has proven to be challenging. The resulting learning algorithm exhibits an all-or-nothing pattern, showing no performance degradation as the constraints increase. This contrasts with previous findings that report a clear decline in performance under increasing constraints. A likely explanation lies in the underlying architecture: while I used a transformer model, previous work used a recurrent network core for meta-learning [22]. This example underscores another key practical constraint placed by the base architecture on meta-learning.

In addition, the neural architectures considered so far lack key cognitive components, such as perception, memory, planning, and language understanding. Recent work has demonstrated that meta-learning is conducive to building in some of these components. For instance, Wang

et al. demonstrated that meta-learned models can be trained end-to-end from pixel to action, enabling them to perceive stimuli directly in pixel space rather than operating at a higher level of abstraction [183]. Before that, Santoro et al. had introduced memory-augmented neural networks that store information in memory, such that the stored representation is stable and retrievable element-wise, while maintaining a fixed number of parameters that are independent of memory size [46]. Jensen, Hennequin, and Mattar recently equipped a meta-reinforcement learning agent with the ability to plan, allowing the agent to sample imagined action sequences from its policy while stopping its interactions with the physical world [184]. Meta-learning agents could also be infused with the ability to use language by co-training them to predict representations from natural language task descriptions and programs induced to generate such tasks, as demonstrated by [185]. Future work should consider bringing these different components together into a single unified model.

But does this not mean that we are returning to the original problem of hand-designing components, as faced by traditional modeling frameworks? Yes, to some extent. This is a necessary trade-off, as there is currently no single architecture that supports the key aspects of cognition discussed above. Nevertheless, I would prefer some—if not all—of these capabilities to emerge through meta-learning, such that the resulting model can capture human behavior across multiple task domains. In a recent work[175], we have taken a step in this direction. We found that an LLM, specifically Llama-3.1 [186], when fine-tuned on the Psych-101 dataset[4], can predict the behavior of held-out participants better than domain-specific cognitive models across multiple tasks and domains. However, even though the cognitive capabilities of Centaur were more aligned with human behavior than those of the base model and other domain-specific models, they still diverged substantially from those of humans [187]. With the availability of large-scale behavioral datasets, such as Psych-101 [175] and PLAICraft [188], future work should investigate whether training the neural architectures discussed so far from scratch on such datasets can (1) help identify architectures that support human-like learning across diverse task settings, and (2) uncover which cognitive capabilities can emerge purely through interaction with an environment, without the need for explicit hand-design.

To explain human learning across multiple time scales, it is necessary to have a model that not only learns within a task but also continues to learn across tasks. However, the objective typically used during meta-learning is only set up to encourage rapid learning within a single task. It enforces the model to learn a newly sampled task in as few trials as possible, but once it is complete, the memory of the task is refreshed, meaning that the model has to learn a new task from scratch again. This issue of forgetting the previous task when solving the present task is known as catastrophic forgetting [189, 190]. It is possible to mitigate this issue and construct meta-continual learning algorithms by (1) episodic recall at the level of trials and tasks using a differential neural dictionary [191] and (2) training a model with an objective that encourages forward and back transfers while still incentivizing it to learn individual tasks in a sample efficient way [181][5]. Future work should consider meta-continual learning algorithms as new frameworks to study learning across multiple time scales, from learning that occurs in an experiment to those that happens

[175]: Binz et al. (2024), 'Centaur: a foundation model of human cognition'

4: Psych-101 is a large-scale dataset containing human behavior around 160 psychological experiments, covering a total of 10,000,000 choices

[189]: Ratcliff (1990), 'Connectionist models of recognition memory: constraints imposed by learning and forgetting functions.'

[190]: McCloskey et al. (1989), 'Catastrophic interference in connectionist networks: The sequential learning problem'

5: In [192], Irie and Lake discuss additional modifications to the meta-learning objective that allow neural networks to display human-like systematic generalization, few-shot learning, and multistep reasoning.

over the course of development.

Finally, the constraints imposed during meta-learning have mostly been on the weights and activations of rate-based neural networks. For instance, I controlled the bits of information needed to store the model using the minimum description length principle in Section 2.3. At the same time, [23] controlled the memory capacity of the network by limiting the number of neurons. Going beyond these, future work should consider richer forms of constraints, from metabolic efficiency to agentic lifetime. For example, the energy- and processing-efficiency of the brain can be captured by imposing sparsity constraints on both the activations (memory) and the parameters (processing) of the neural network, as it has been shown to result in substantial gains in efficiency with minimal loss in performance [193]. Neural architectures with implicit constraints, like a recurrent network of spiking neurons shown to possess long-term memory and the capacity to learn-to-learn [194], offer a more biologically plausible approach to implement bounded meta-learned inference. Furthermore, there are other constraints, such as ecological uncertainty, task horizon, thinking time, and agent lifetime, that can be introduced within the meta-learning setup [195].

In general, understanding the downstream effects of architectural and objective-based constraints is a crucial next step in developing more plausible meta-learned models of human learning and decision-making.

Bridging different levels of analysis

Human learning and the factors driving it can be understood at different levels of analysis, as proposed by Marr[6]. Traditionally, modeling frameworks such as connectionist and Bayesian models commit to explaining human learning at a specific level of analysis. For instance, connectionist models excel at providing explanations at the algorithmic and implementation levels, and Bayesian models are successful at explaining behavior at the computational and algorithmic levels (with appropriate approximations [93]). By allowing us to derive neural networks that can implement approximation to Bayes-optimal policies, meta-learning provides a unifying framework to bridge Marr's three levels of analysis.

However, this key strength of meta-learning has yet to be fully utilized. Previous work, especially in neuroscience, has mostly focused on understanding how meta-learning shapes neural representations such that it allows rapid learning [47, 197]. For instance, Hattori et al. have shown how meta-learning shapes population value coding in the orbito-frontal cortex, across multiple sessions, to guide the trial-by-trial behavior of the mice in a reversal learning task [197].

Future work should leverage this strength to understand the factors that drive human learning at multiple levels. For example, it has recently been shown that discrete symbolic priors from Bayesian models [140] can be distilled into the weights and activations of the neural network. Taking a step further, one can identify the population of in silico neurons in meta-learned models that code symbolic priors, using tools such as sparse autoencoders [7], and build encoding models from their activity to activity patterns in the brain, particularly the prefrontal cortex (PFC) [201].

6: David Marr [196] presented a framework that can be used to study any information processing system, including the human brain. He proposed that any information processing system systems can be understood at three levels of abstraction: (1) computational level; (2) algorithmic level; (3) implementation level. The first level focuses on characterizing the computational problem that is being solved, the second on the representational structure and algorithm used to solve the computational problem, and the third on the hardware on which the algorithm is implemented.

7: In a work not included in this thesis, we have shown that sparse autoencoders [198, 199] that can be used to identify representations that closely match temporal difference (TD) errors in LLMs during in-context learning [200]

This would potentially allow us to understand how the brain supports and utilizes abstract structures for learning and generalization.

Extending to other animal species

So far, I have focused mainly on testing theories of *human* learning with meta-learning. I believe that the benefits of meta-learning become even more pronounced when used to investigate learning in other animal species. This is because animal studies allow precise control over the rearing environments, the type of rewards that are provided and their curriculum structure, and perhaps even genetic predisposition of the animal under study. Of particular interest for meta-learning would be controlled-rearing experiments, where the environment in which the animal is raised is experimentally controlled, and all exhibited behavior is tracked through automated systems. For instance, Wood and colleagues [202] have successfully conducted a virtual reality-based controlled rearing experiment in newborn chicks to investigate how visual experience influences high-level cognitive abilities, such as object recognition. Future work should build on this research by first recording all sensory experiences of a chick, along with the behaviors it exhibits. Then, training a neural network from scratch on the collected dataset using a meta-learning objective. This would potentially take us a step closer to settling the nativism versus constructivism debate, that is, determining which capabilities are acquired through experience and which might have been built into us by evolution.

3.3 Implications

Insights drawn from testing theories of human learning using meta-learning can not only be helpful for the field of cognitive psychology, but also be beneficial to other fields such as developmental psychology, computational psychiatry, educational psychology, and artificial intelligence.

Developmental psychologists have shown great interest in understanding how environmental factors shape learning in the developing mind and how its capacities mature with experience [203–205]. However, prior modeling efforts have primarily focused on learning within shorter time scales, like learning within the duration of a psychological experiment. Meta-learning, on the other hand, enables the investigation of how environmental factors shape human learning and development across multiple nested time scales [21]. For example, they can be used to simulate the effects of volatility of rewards, controllability of actions, and richness of surroundings over the course of development on downstream behavior. The findings from such an effort can be used to design nourishing environments for children that enable them to reach their full potential.

The modeling pipeline described above can also be extended to investigate the risks posed by adverse environments to mental well-being. Research in psychiatry has demonstrated that early life stress can have profound negative implications, such as disrupting reward-guided learning and

decision-making, which can persist into adulthood [206, 207]. Computational psychiatry has attempted to formalize these experimental findings into computational theories, but the resulting theories are mainly limited to shorter time scales [208, 209]. Meta-learning remedies this issue by allowing researchers to investigate the side effects of negative childhood experiences computationally. For instance, it can be used to study whether sparse rewarding environments growing up lead to reward insensitivity at the core of depression; (2) volatile environments result in anxiety as a coping mechanism; (3) environments with limited agency result in learned helplessness or obsessive impulsivity. The insights drawn from such a study can potentially be used to create new psychotherapies to overcome learned maladaptive behavior.

Educational psychologists have devoted considerable resources to designing learning curricula and incentive structures that aid in effective skill development in children and adults alike [210, 211]. Theories formulated through this research have been instrumental in the development of educational programs around the world [212]. However, once a curriculum is established, enhancing it becomes significantly challenging. This difficulty largely stems from the lengthy process required to test and validate the effectiveness of a revised curriculum [213, 214]. Meta-learning can potentially accelerate this process, as it enables in-silico testing and iteration of learning curricula. Additionally, the resulting teaching curricula can be customized to maximize immediate task performance or long-term retention through an objective-driven optimization procedure.

LLMs have been at the center of recent success in AI [215]. The LLM training objective is a special case of the meta-learning objective shown in Equation 1.3, with the meta-learning distribution replaced by human language data scraped from the Internet [216]. LLMs, despite their fragility, are extremely powerful models. They, like meta-learned models, can quickly learn a new task based on examples provided within their context window, without requiring any updates to the network weights [48]. They can serve as good base models for solving a variety of different tasks, e.g., predicting human behavior [175], and for acquiring additional capabilities, e.g., reasoning [217]. Recently, I have demonstrated that they can even discover cognitive models that are superior to those proposed in the literature [218]. Given the strong push to expand existing AI systems by training increasingly large neural networks on richer datasets that include audio, video, and actions, we are on the verge of understanding the boundaries of meta-learning and its utility in building general learning systems.

[218]: Rmus et al. (2025), 'Generating Computational Cognitive Models using Large Language Models'

3.4 Conclusion

Human learning is a remarkable feat. It renders challenging tasks, from integrating to a new country to mastering a new sport, so deceptively simple that the astonishing complexity underlying it remains masked. Over the years, cognitive scientists have developed a wide range of different frameworks to understand the factors that drive human learning. Some of these frameworks have emphasized the interplay between agents' goals and environmental structure, while others have focused

on mechanistic constraints. Yet, no single framework has succeeded in offering a complete account of human learning.

To fill this gap, I introduced a framework based on meta-learning that simultaneously considers goals, environments, and resource constraints, while allowing the construction of models that learn to learn instead of being tailored to individual tasks. I began by illustrating how meta-learning facilitates deriving Bayes-optimal learning algorithms, discussed how it combines the strengths from Bayesian and connectionist models of cognition, and highlighted its connection to well-established frameworks such as rational analysis. Subsequently, I showcased the unique capabilities of meta-learning through four different studies. In the first, I used meta-learning to conduct rational analysis of optimism bias, showing that it can viewed as a rational strategy in light of the environments the agent has previously interacted with. In the second study, I demonstrated that resource-rational models derived using meta-learning capture zero-shot compositional inference in humans. In the third study, I constructed models of ecological rationality using meta-learning to demonstrate how they can capture various aspects of human category learning. In the final study, I extended the idea further, showing that meta-learning on ecologically valid environments is sufficient to account for a substantial portion of human behavior.

In an extended discussion, I related these findings to three classical frameworks for studying human cognition: rational analysis, resource-rational analysis, and ecological rationality, and emphasized the limitations of traditional modeling frameworks addressed by my approach. In addition, I highlighted the current limitations faced by meta-learned models and the possible future directions. Specifically, I recommended that future studies should: (1) scale meta-learning to richer tasks, facilitating explanations of naturalistic decision-making; (2) incorporate more plausible architectures, objectives, and constraints to better account for resource-limited cognition; (3) ground learned representations in neural mechanisms, thus bridging Marr's levels of analysis; and (4) extend research to other animal species, particularly via controlled-rearing experiments, as they allow for the causal evaluation of predictions derived from meta-learned models.

Overall, my findings suggest that meta-learning offers a domain-general framework for building models of human learning. I intend to continue advancing the research program we have developed around meta-learning, in the hope that it ultimately fulfills its promise as"a general theory of natural intelligence that is—more than [its] classical counterpart—suitable for the real world." [219].

Bibliography

Here are the references in citation order.

[1] Brenden M Lake et al. 'Building machines that learn and think like people'. In: *Behavioral and brain sciences* 40 (2017) (cited on page 3).

[2] Susan Carey and Elsa Bartlett. 'Acquiring a single new word.' In: *Reports on Child Language Development* (1978) (cited on page 3).

[3] Klaus M Stiefel and Jay S Coggan. 'The energy challenges of artificial superintelligence'. In: *Frontiers in Artificial Intelligence* 6 (2023), p. 1240653 (cited on page 3).

[4] Alison Gopnik et al. 'A theory of causal learning in children: causal maps and Bayes nets.' In: *Psychological review* 111.1 (2004), p. 3 (cited on page 3).

[5] Kelsey R Allen, Kevin A Smith, and Joshua B Tenenbaum. 'Rapid trial-and-error learning with simulation supports flexible tool use and physical reasoning'. In: *Proceedings of the National Academy of Sciences* 117.47 (2020), pp. 29302–29310 (cited on page 3).

[6] John R Anderson. 'Is human cognition adaptive?' In: *Behavioral and brain sciences* 14.3 (1991), pp. 471–485 (cited on pages 3, 8, 10, 12, 208).

[7] Falk Lieder and Thomas L Griffiths. 'Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources'. In: *Behavioral and Brain Sciences* 43 (2020) (cited on pages 3, 11, 209).

[8] Peter M Todd and Gerd Gigerenzer. *Ecological rationality: Intelligence in the world*. Cary, NC: Oxford University Press, 2012 (cited on pages 3, 13, 14, 139).

[9] Tom M Mitchell. *Machine learning*. Vol. 1. 9. 1997 (cited on page 3).

[10] Y Bengio, S Bengio, and J Cloutier. 'Learning a synaptic learning rule'. In: *IJCNN-91-Seattle International Joint Conference on Neural Networks*. Vol. 2. IEEE. 1991, 969–vol (cited on page 3).

[11] Jürgen Schmidhuber. 'Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook'. PhD thesis. Technische Universität München, 1987 (cited on page 3).

[12] Sebastian Thrun and Lorien Pratt. 'Learning to learn: Introduction and overview'. In: *Learning to learn*. Springer, 1998, pp. 3–17 (cited on page 3).

[13] F. Gregory Ashby and W. Todd Maddox. 'Human Category Learning'. en. In: *Annual Review of Psychology* 56.1 (Feb. 2005), pp. 149–178. DOI: 10.1146/annurev.psych.56.091103.070217. (Visited on 01/24/2024) (cited on page 4).

[14] Roger N Shepard, Carl I Hovland, and Herbert M Jenkins. 'Learning and memorization of classifications'. In: *Psychological Monographs: General and Applied* 75.13 (1961), pp. 1–42. DOI: 10.1037/h0093825 (cited on pages 4, 139, 212).

[15] Jerome R Busemeyer et al. 'Learning functional relations based on experience with input–output pairs by humans and artificial neural networks'. In: *Knowledge concepts and categories*. Psychology Press, 2013, pp. 405–437 (cited on page 4).

[16] Gerd Gigerenzer and Wolfgang Gaissmaier. 'Heuristic decision making'. In: *Annual review of psychology* 62 (2011), pp. 451–482 (cited on pages 4, 14).

[17] Matthew Botvinick et al. 'Reinforcement learning, fast and slow'. In: *Trends in cognitive sciences* 23.5 (2019), pp. 408–422 (cited on page 4).

[18] Thad A Polk and Colleen M Seifert. *Cognitive modeling*. MIT Press, 2002 (cited on page 4).

[19] Ron Sun. *The Cambridge handbook of computational cognitive sciences*. Cambridge University Press, 2023 (cited on page 4).

[20] Jane X Wang. 'Meta-learning in natural and artificial intelligence'. In: *arXiv [cs.AI]* (Nov. 2020) (cited on page 4).

[21] Kate Nussenbaum and Catherine A Hartley. 'Understanding the development of reward learning through the lens of meta-learning'. In: *Nature Reviews Psychology* (2024), pp. 1–15 (cited on pages 4, 218).

[22] Marcel Binz et al. 'Heuristics From Bounded Meta-Learned Inference'. In: *Psychological review* (2022) (cited on pages 4, 7, 168, 212, 214, 215).

[23] Ishita Dasgupta et al. 'A theory of learning to infer.' In: *Psychological review* 127.3 (2020), p. 412 (cited on pages 4, 217).

[24] Sreejan Kumar et al. 'Meta-Learning of Structured Task Distributions in Humans and Machines'. In: *International Conference on Learning Representations*. 2021 (cited on page 4).

[25] Jonathan Baxter. 'Theoretical models of learning to learn'. In: *Learning to learn*. Springer, 1998, pp. 71–94 (cited on page 4).

[26] Erin Grant et al. 'Recasting gradient-based meta-learning as hierarchical bayes'. In: *6th International Conference on Learning Representations, ICLR 2018*. 2018 (cited on pages 4, 6).

[27] Pengyu Yuan and Hien Van Nguyen. 'Model-based meta learning'. In: *Meta Learning With Medical Imaging and Health Informatics Applications*. Elsevier, 2023, pp. 65–74 (cited on page 4).

[28] Mike Huisman, Jan N Van Rijn, and Aske Plaat. 'A survey of deep meta-learning'. In: *Artificial Intelligence Review* 54.6 (2021), pp. 4483–4541 (cited on page 4).

[29] Timothy Hospedales et al. 'Meta-learning in neural networks: A survey'. In: *IEEE transactions on pattern analysis and machine intelligence* 44.9 (2021), pp. 5149–5169 (cited on page 4).

[30] Anna Vettoruzzo et al. 'Advances and challenges in meta-learning: A technical review'. In: *IEEE transactions on pattern analysis and machine intelligence* 46.7 (2024), pp. 4763–4779 (cited on pages 4–6).

[31] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 'Model-agnostic meta-learning for fast adaptation of deep networks'. In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1126–1135 (cited on page 5).

[32] Sebastian Ruder. 'An overview of gradient descent optimization algorithms'. In: *arXiv preprint arXiv:1609.04747* (2016) (cited on page 5).

[33] Zhenguo Li et al. 'Meta-sgd: Learning to learn quickly for few-shot learning'. In: *arXiv preprint arXiv:1707.09835* (2017) (cited on page 5).

[34] Chelsea Finn, Kelvin Xu, and Sergey Levine. 'Probabilistic model-agnostic meta-learning'. In: *Advances in neural information processing systems* 31 (2018) (cited on page 5).

[35] Alex Nichol, Joshua Achiam, and John Schulman. 'On first-order meta-learning algorithms'. In: *arXiv preprint arXiv:1803.02999* (2018) (cited on pages 5, 6).

[36] Andrew M Saxe, James L McClelland, and Surya Ganguli. 'Exact solutions to the nonlinear dynamics of learning in deep linear neural networks'. In: *arXiv preprint arXiv:1312.6120* (2013) (cited on page 6).

[37] Sebastian Goldt et al. 'Dynamics of stochastic gradient descent for two-layer neural networks in the teacher-student setup'. In: *Advances in neural information processing systems* 32 (2019) (cited on page 6).

[38] Yao Zhang et al. 'Energy–entropy competition and the effectiveness of stochastic gradient descent in machine learning'. In: *Molecular Physics* 116.21-22 (2018), pp. 3214–3223 (cited on page 6).

[39] R Thomas McCoy et al. 'Universal linguistic inductive biases via meta-learning'. In: *arXiv preprint arXiv:2006.16324* (2020) (cited on page 6).

[40] Rachit Dubey et al. 'Connecting Context-specific Adaptation in Humans to Meta-learning'. In: *arXiv preprint arXiv:2011.13782* (2020) (cited on page 6).

[41] Nicolas Zucchet et al. 'A contrastive rule for meta-learning'. In: *Advances in neural information processing systems* 35 (2022), pp. 25921–25936 (cited on page 6).

[42] Simon Guiroy, Vikas Verma, and Christopher Pal. 'Towards understanding generalization in gradient-based meta-learning'. In: *arXiv preprint arXiv:1907.07287* (2019) (cited on page 6).

[43] Sepp Hochreiter, A. Steven Younger, and Peter R. Conwell. 'Learning to learn using gradient descent'. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2130 (2001), pp. 87–94. DOI: 10.1007/3-540-44668-0{_}13 (cited on page 6).

[44] Ashish Vaswani et al. 'Attention is all you need'. In: *Advances in neural information processing systems* 30 (2017) (cited on pages 6, 209, 215).

[45] Albert Gu, Karan Goel, and Christopher Ré. 'Efficiently modeling long sequences with structured state spaces'. In: *arXiv preprint arXiv:2111.00396* (2021) (cited on pages 6, 215).

[46] Adam Santoro et al. 'Meta-learning with memory-augmented neural networks'. In: *International conference on machine learning*. PMLR. 2016, pp. 1842–1850 (cited on pages 6, 216).

[47] Jane X Wang et al. 'Learning to reinforcement learn'. In: *arXiv preprint arXiv:1611.05763* (2016) (cited on pages 6, 101, 217).

[48] Tom Brown et al. 'Language models are few-shot learners'. In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901 (cited on pages 6, 214, 215, 219).

[49] Mark Sandler et al. 'Meta-learning bidirectional update rules'. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 9288–9300 (cited on page 7).

[50] A Paszke. 'Pytorch: An imperative style, high-performance deep learning library'. In: *arXiv preprint arXiv:1912.01703* (2019) (cited on page 7).

[51] Pedro A Ortega et al. 'Meta-learning of sequential strategies'. In: *arXiv preprint arXiv:1905.03030* (2019) (cited on pages 7, 11, 21).

[52] John R Anderson. *The adaptive character of thought*. Hillsdale, NJ: Erlbaum, 1990 (cited on pages 7, 9, 168).

[53] Marcel Binz and Eric Schulz. 'Modeling Human Exploration Through Resource-Rational Reinforcement Learning'. In: *Advances in Neural Information Processing Systems*. Oct. 2022. (Visited on 06/07/2023) (cited on pages 7, 209, 214).

[54] Marvin L Minsky. 'Logical versus analogical or symbolic versus connectionist or neat versus scruffy'. In: *AI magazine* 12.2 (1991), pp. 34–34 (cited on page 7).

[55] David Willshaw. 'Symbolic and Subsymbolic Approaches to Cognition'. In: *Neural Computation and Psychology: Proceedings of the 3rd Neural Computation and Psychology Workshop (NCPW3), Stirling, Scotland, 31 August–2 September 1994*. Springer. 1995, pp. 3–18 (cited on page 7).

[56] Peter Markie and Marina Folescu. 'Rationalism vs. empiricism'. In: (2004) (cited on page 7).

[57] Thomas L Griffiths et al. 'Probabilistic models of cognition: Exploring representations and inductive biases'. In: *Trends in cognitive sciences* 14.8 (2010), pp. 357–364 (cited on pages 7, 9, 207).

[58] James L McClelland et al. 'Letting structure emerge: connectionist and dynamical systems approaches to cognition'. In: *Trends in cognitive sciences* 14.8 (2010), pp. 348–356 (cited on page 8).

[59] Thomas Griffiths, Charles Kemp, and Joshua B Tenenbaum. *Bayesian models of cognition*. Carnegie Mellon University, 2008 (cited on page 8).

[60] Thomas L Griffiths, Nick Chater, and Joshua B Tenenbaum. *Bayesian models of cognition: reverse engineering the mind*. MIT Press, 2024 (cited on pages 8, 21).

[61] Christopher G Lucas et al. 'A rational model of function learning'. In: *Psychonomic bulletin & review* 22.5 (2015), pp. 1193–1215 (cited on pages 8, 9, 168, 207, 208).

[62] Ishita Dasgupta et al. 'Causal reasoning from meta-reinforcement learning'. In: *arXiv preprint arXiv:1901.08162* (2019) (cited on page 8).

[63] John R Anderson. 'The adaptive nature of human categorization'. In: *Psychol. Rev.* 98.3 (July 1991), pp. 409–429. DOI: 10.1037/0033-295X.98.3.409 (cited on page 8).

[64] Michael SC Thomas and James L McClelland. 'Connectionist models of cognition'. In: *The Cambridge handbook of computational psychology* (2008), pp. 23–58 (cited on pages 8, 21).

[65] Kazuki Irie and Brenden M Lake. 'Neural networks that overcome classic challenges through practice'. In: *arXiv [cs.AI]* (Oct. 2024) (cited on page 8).

[66] James L McClelland. *Parallel distributed processing: Implications for cognition and development*. Depts. of Computer Science and Psychology, Carnegie Mellon University, 1988 (cited on page 8).

[67] Andrew M Saxe, James L McClelland, and Surya Ganguli. 'A mathematical theory of semantic development in deep neural networks'. In: *Proceedings of the National Academy of Sciences* 116.23 (2019), pp. 11537–11546 (cited on page 8).

[68] John Kruschke. 'ALCOVE: A connectionist model of human category learning'. In: *Advances in Neural Information Processing Systems* 3 (1990) (cited on page 8).

[69] Martha J Farah and James L McClelland. 'A computational model of semantic memory impairment: Modality specificity and emergent category specificity.' In: *Journal of Experimental Psychology: General* 120.4 (1991), p. 339 (cited on page 8).

[70] David E Rumelhart and James L McClelland. 'On learning the past tenses of English verbs'. In: *Psycholinguistics: critical concepts in psychology* 4 (1986), pp. 216–271 (cited on page 8).

[71] John K. Kruschke. 'ALCOVE: An exemplar-based connectionist model of category learning.' In: *Psychological Review* 99.1 (May 1992). Publisher: US: American Psychological Association, p. 22. DOI: 10.1037/0033-295X.99.1.22. (Visited on 08/31/2021) (cited on page 8).

[72] Marcel Binz et al. *Meta-Learned Models of Cognition*. Apr. 2023. (Visited on 05/08/2023) (cited on page 8).

[73] Desmond C Ong et al. 'Probabilistic programming versus meta-learning as models of cognition'. In: *The Behavioral and brain sciences* 47 (2024), e158 (cited on page 9).

[74] R Thomas McCoy and Thomas L Griffiths. 'Meta-learning as a bridge between neural networks and symbolic Bayesian models.' In: *Behavioral & Brain Sciences* 47 (2024) (cited on page 9).

[75] Geoffrey E Hinton and James A Anderson. 'Parallel models of associative memory'. In: (1989) (cited on page 9).

[76] Reuben Feinman and Brenden M Lake. 'Learning task-general representations with generative neuro-symbolic modeling'. In: *arXiv preprint arXiv:2006.14448* (2020) (cited on page 9).

[77] Yanli Zhou, Reuben Feinman, and Brenden M Lake. 'Compositional diversity in visual concept learning'. In: *Cognition* 244 (2024), p. 105711 (cited on page 9).

[78] Kevin Ellis et al. 'DreamCoder: growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning'. In: *Philosophical Transactions of the Royal Society A* 381.2251 (2023), p. 20220050 (cited on page 9).

[79] Kevin Ellis. 'Human-like Few-Shot Learning via Bayesian Reasoning over Natural Language'. In: (Nov. 2023) (cited on page 9).

[80] Jane X Wang et al. 'Prefrontal cortex as a meta-reinforcement learning system'. In: *Nature neuroscience* 21.6 (2018), pp. 860–868 (cited on page 9).

[81] Nick Chater and Mike Oaksford. 'Ten Years of the Rational Analysis of Cognition'. In: *Trends in Cognitive Sciences* 3.2 (Feb. 1999), pp. 57–65. DOI: 10.1016/S1364-6613(98)01273-X. (Visited on 06/03/2023) (cited on pages 10, 13).

[82] Mike Oaksford and Nick Chater. 'A rational analysis of the selection task as optimal data selection.' In: *Psychological review* 101.4 (1994), p. 608 (cited on page 10).

[83] Thomas L Griffiths and Joshua B Tenenbaum. 'Optimal predictions in everyday cognition'. In: *Psychological science* 17.9 (2006), pp. 767–773 (cited on page 10).

[84] Mike Oaksford, Nick Chater, et al. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press, 2007 (cited on page 10).

[85] Iris Van Rooij. 'The tractable cognition thesis'. In: *Cognitive science* 32.6 (2008), pp. 939–984 (cited on page 10).

[86] Nick Chater and Mike Oaksford. 'The Rational Analysis of Mind and Behavior'. In: *Synthese* 122.1/2 (2000), pp. 93–131 (cited on page 10).

[87] Valerie M Chase, Ralph Hertwig, and Gerd Gigerenzer. 'Visions of rationality'. In: *Trends in cognitive sciences* 2.6 (1998), pp. 206–214 (cited on pages 11, 14).

[88] Tali Sharot. *The Optimism Bias: Why We're Wired to Look on the Bright Side*. Hachette UK, Jan. 2012 (cited on page 11).

[89] Stefano Palminteri and Maël Lebreton. 'The Computational Roots of Positivity and Confirmation Biases in Reinforcement Learning'. In: *Trends in Cognitive Sciences* 26.7 (July 2022), pp. 607–621. DOI: 10.1016/j.tics.2022.04.005. (Visited on 10/05/2022) (cited on page 11).

[90] Eric Joel Horvitz. *Computation and action under bounded resources*. stanford university, 1991 (cited on page 11).

[91] Stuart J Russell and Devika Subramanian. 'Provably bounded-optimal agents'. In: *Journal of Artificial Intelligence Research* 2 (1994), pp. 575–609 (cited on page 11).

[92] Richard L Lewis, Andrew Howes, and Satinder Singh. 'Computational rationality: Linking mechanism and behavior through bounded utility maximization'. In: *Topics in cognitive science* 6.2 (2014), pp. 279–311 (cited on pages 11, 213).

[93] Thomas L Griffiths, Falk Lieder, and Noah D Goodman. 'Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic'. In: *Topics in cognitive science* 7.2 (2015), pp. 217–229 (cited on pages 11, 12, 217).

[94] F Gobet et al. 'Chunking mechanisms in human learning'. en. In: *Trends Cogn. Sci.* 5.6 (June 2001), pp. 236–243. DOI: 10.1016/s1364-6613(00)01662-4 (cited on page 12).

[95] Shuchen Wu et al. 'Building, Reusing, and Generalizing Abstract Representations from Concrete Sequences'. In: *arXiv preprint arXiv:2410.21332* (2024) (cited on page 12).

[96] Momchil S Tomov et al. 'Discovery of hierarchical representations for efficient planning'. In: *PLoS computational biology* 16.4 (2020), e1007594 (cited on page 12).

[97] H A Simon. 'Rational choice and the structure of the environment'. en. In: *Psychol. Rev.* 63.2 (Mar. 1956), pp. 129–138. DOI: 10.1037/h0042769 (cited on pages 12, 14).

[98] Xavier Gabaix. 'A sparsity-based model of bounded rationality'. en. In: *Q. J. Econ.* 129.4 (Nov. 2014), pp. 1661–1710. DOI: 10.1093/qje/qju024 (cited on page 12).

[99] Robert L Goldstone. 'Perceptual learning'. In: *Annual review of psychology* 49.1 (1998), pp. 585–612 (cited on page 12).

[100] Samuel Planton et al. 'A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans'. In: *PLoS computational biology* 17.1 (2021), e1008598 (cited on page 12).

[101] Timothy F Brady, Talia Konkle, and George A Alvarez. 'Compression in visual working memory: using statistical regularities to form more efficient memory representations.' In: *Journal of Experimental Psychology: General* 138.4 (2009), p. 487 (cited on page 12).

[102] Willem B Verwey. 'Buffer loading and chunking in sequential keypressing.' In: *Journal of Experimental Psychology: Human Perception and Performance* 22.3 (1996), p. 544 (cited on page 12).

[103] Falk Lieder, Thomas L Griffiths, and Ming Hsu. 'Overrepresentation of extreme events in decision making reflects rational use of cognitive resources'. en. In: *Psychol. Rev.* 125.1 (Jan. 2018), pp. 1–32. DOI: 10.1037/rev0000074 (cited on page 12).

[104] Xavier Gabaix and David Laibson. 'Bounded rationality and directed cognition'. In: *Harvard University* 20 (2005) (cited on page 12).

[105] Adam N Sanborn, Thomas L Griffiths, and Daniel J Navarro. 'A more rational model of categorization'. In: *Proceedings of the 28th annual conference of the cognitive science society*. Mahwah, NJ. 2006, pp. 726–731 (cited on page 12).

[106] Adam N Sanborn, Thomas L Griffiths, and Daniel J Navarro. 'Rational approximations to rational models: alternative algorithms for category learning.' In: *Psychological review* 117.4 (2010), p. 1144 (cited on page 12).

[107] Adam N Sanborn. 'Types of approximation for probabilistic cognition: Sampling and variational'. In: *Brain and cognition* 112 (2017), pp. 98–101 (cited on page 13).

[108] Herbert A Simon. 'Invariants of human behavior'. In: *Annual review of psychology* 41.1 (1990), pp. 1–20 (cited on page 13).

[109] Gerd Gigerenzer, Peter M Todd, the ABC Research Group, et al. *Simple heuristics that make us smart*. Oxford University Press, 2000 (cited on pages 13, 14).

[110] Ralph Hertwig et al. 'Studies in ecological rationality'. In: *Topics in Cognitive Science* 14.3 (2022), pp. 467–491 (cited on pages 13, 211).

[111] Gerd Gigerenzer and Daniel G Goldstein. 'Reasoning the fast and frugal way: models of bounded rationality.' In: *Psychological review* 103.4 (1996), p. 650 (cited on pages 14, 168).

[112] Daniel G Goldstein and Gerd Gigerenzer. 'Models of ecological rationality: the recognition heuristic.' In: *Psychological review* 109.1 (2002), p. 75 (cited on pages 14, 168).

[113] Gerd Gigerenzer, Jean Czerlinski, and Laura Martignon. 'How good are fast and frugal heuristics?' In: *Decision science and technology: Reflections on the contributions of Ward Edwards*. Springer, 1999, pp. 81–103 (cited on page 14).

[114] Nick Chater et al. 'Fast, frugal, and rational: How rational norms explain behavior'. In: *Organizational behavior and human decision processes* 90.1 (2003), pp. 63–86 (cited on page 14).

[115] Jean Czerlinski, Gerd Gigerenzer, Daniel G Goldstein, et al. 'How good are simple heuristics'. In: *Simple heuristics that make us smart* (1999), pp. 97–118 (cited on page 14).

[116] Leonard J Savage. *The foundations of statistics*. Courier Corporation, 1972 (cited on page 14).

[117] Peter M Todd and Gerd Gigerenzer. 'Environments that make us smart: Ecological rationality'. In: *Current directions in psychological science* 16.3 (2007), pp. 167–171 (cited on page 14).

[118] Jerome H Barkow, Leda Cosmides, and John Tooby. *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press, 1995 (cited on page 14).

[119] John W Payne, James R Bettman, and Eric J Johnson. *The adaptive decision maker*. Cambridge university press, 1993 (cited on page 14).

[120] Gerd Gigerenzer, Peter M. Todd, and the ABC Research Group. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press, 1999 (cited on page 14).

[121] Liz Allen et al. 'Publishing: Credit where credit is due'. In: *Nature* 508.7496 (2014), pp. 312–313 (cited on page 15).

[122] Copyright Clearance Center. *RightsLink® Permissions Portal*. https://s100.copyright.com/AppDispatchServlet#formTop. Accessed: 2025-06-29. 2025 (cited on page 22).

[123] Germain Lefebvre et al. 'Behavioural and Neural Characterization of Optimistic Reinforcement Learning'. In: *Nature Human Behaviour* 1.4 (Mar. 2017), p. 0067. DOI: 10.1038/s41562-017-0067. (Visited on 01/16/2023) (cited on page 81).

[124] Valérian Chambon et al. 'Information about Action Outcomes Differentially Affects Learning from Self-Determined versus Imposed Choices'. In: *Nature Human Behaviour* 4.10 (Oct. 2020), pp. 1067–1079. DOI: 10.1038/s41562-020-0919-5. (Visited on 06/25/2023) (cited on pages 81, 210).

[125] Germain Lefebvre, Christopher Summerfield, and Rafal Bogacz. 'A Normative Account of Confirmation Bias During Reinforcement Learning'. In: *Neural Computation* 34.2 (Jan. 2022), pp. 307–337. DOI: 10.1162/neco_a_01455. (Visited on 11/08/2022) (cited on pages 81, 210).

[126] Stefano Palminteri and Maël Lebreton. 'The computational roots of positivity and confirmation biases in reinforcement learning'. In: *Trends in Cognitive Sciences* 26.7 (2022), pp. 607–621 (cited on page 81).

[127] International Conference on Machine Learning. *ICML Copyright FAQ*. Accessed: 2025-06-27. 2025. URL: https://icml.cc/FAQ/Copyright (cited on pages 81, 140).

[128] Yan Duan et al. 'Rl $\hat{2}$: Fast reinforcement learning via slow reinforcement learning'. In: *arXiv preprint arXiv:1611.02779* (2016) (cited on page 101).

[129] Marcel Binz and Eric Schulz. 'Modeling human exploration through resource-rational reinforcement learning'. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 31755–31768 (cited on page 101).

[130] Roger G Barker. *Ecological psychology: Concepts and Methods for Studying the Environment of Human Behavior*. Stanford, CA: Stanford University Press, 1968 (cited on page 139).

[131] Kenneth R Hammond. *Ecological validity: Then and now*. 1998 (cited on page 139).

[132] Bernd Bischl et al. 'OpenML Benchmarking Suites'. In: *arXiv:1708.03731v2 [stat.ML]* (2019) (cited on page 139).

[133] J David Smith and John Paul Minda. 'Prototypes in the mist: The early epochs of category learning.' In: *Journal of Experimental Psychology: Learning, memory, and cognition* 24.6 (1998), p. 1411 (cited on pages 139, 212).

[134] Egon Brunswik. 'Representative design and probabilistic theory in a functional psychology.' In: *Psychological review* 62.3 (1955), p. 193 (cited on page 168).

[135] Arjun Devraj, Qiong Zhang, and Tom Griffiths. 'The dynamics of exemplar and prototype representations depend on environmental statistics'. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 43. 2021 (cited on pages 168, 212).

[136] Stephen P Badham, Adam N Sanborn, and Elizabeth A Maylor. 'Deficits in category learning in older adults: Rule-based versus clustering accounts'. en. In: *Psychol. Aging* 32.5 (Aug. 2017), pp. 473–488. DOI: 10.1037/pag0000183 (cited on pages 168, 208, 212).

[137] Brenden M Lake and Steven T Piantadosi. 'People infer recursive visual concepts from just a few examples'. In: *Computational Brain & Behavior* 3.1 (2020), pp. 54–65 (cited on page 207).

[138] Xi Chen et al. 'PaLI: A Jointly-Scaled Multilingual Language-Image Model'. In: *The Eleventh International Conference on Learning Representations* (cited on page 207).

[139] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. 'Human-level concept learning through probabilistic program induction'. In: *Science* 350.6266 (2015), pp. 1332–1338 (cited on pages 207, 208).

[140] Ioana Marinescu, R Thomas McCoy, and Thomas L Griffiths. 'Distilling Symbolic Priors for Concept Learning into Neural Networks'. In: *arXiv preprint arXiv:2402.07035* (2024) (cited on pages 207, 217).

[141] Daniel R Little, Richard M Shiffrin, and Simon M Laham. 'Function estimation: Quantifying individual differences of hand-drawn functions'. In: *Memory & Cognition* (2024), pp. 1–20 (cited on page 208).

[142] Woo-kyoung Ahn, William F Brewer, and Raymond J Mooney. 'Schema acquisition from a single example.' In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18.2 (1992), p. 391 (cited on page 208).

[143] Jerry A Fodor and Zenon W Pylyshyn. 'Connectionism and cognitive architecture: A critical analysis'. In: *Cognition* 28.1-2 (1988), pp. 3–71 (cited on page 208).

[144] Jordan B Pollack. 'Recursive distributed representations'. In: *Artificial Intelligence* 46.1-2 (1990), pp. 77–105 (cited on page 208).

[145] Paul Smolensky. 'Tensor product variable binding and the representation of symbolic structures in connectionist systems'. In: *Artificial intelligence* 46.1-2 (1990), pp. 159–216 (cited on page 208).

[146] Trenton Kriete et al. 'Indirection and symbol-like processing in the prefrontal cortex and basal ganglia'. In: *Proceedings of the National Academy of Sciences* 110.41 (2013), pp. 16390–16395 (cited on page 208).

[147] Randall C O'Reilly et al. 'How limited systematicity emerges: A computational cognitive neuroscience approach'. In: *The architecture of cognition: Rethinking fodor and Pylyshyn's systematicity challenge* (2013) (cited on page 209).

[148] Kent Johnson. 'On the systematicity of language and thought'. In: *The Journal of Philosophy* 101.3 (2004), pp. 111–139 (cited on page 209).

[149] Brenden M Lake and Marco Baroni. 'Human-like systematic generalization through a meta-learning neural network'. In: *Nature* (2023), pp. 1–7 (cited on page 209).

[150] Tenenbaum. *Joshua Tenenbaum's homepage*. http://web.mit.edu/cocosci/josh.html. [Online; accessed 9-November-2021]. 2021 (cited on pages 209, 212).

[151] Romain D Cazé and Matthijs AA van der Meer. 'Adaptive properties of differential learning rates for positive and negative outcomes'. In: *Biological cybernetics* 107.6 (2013), pp. 711–719 (cited on page 210).

[152] Kefan Song et al. 'Reward Is Enough: LLMs Are In-Context Reinforcement Learners'. In: *arXiv preprint arXiv:2506.06303* (2025) (cited on page 210).

[153] AnthropicAI. *Introducing Claude 2.1*. en. https://www.anthropic.com/news/claude-2-1. Accessed: 2024-1-29. 2023 (cited on page 211).

[154] Sepp Hochreiter, A Steven Younger, and Peter R Conwell. 'Learning to learn using gradient descent'. In: *Artificial Neural Networks—ICANN 2001: International Conference Vienna, Austria, August 21–25, 2001 Proceedings 11*. Springer. 2001, pp. 87–94 (cited on page 211).

[155] Vadim Borisov et al. 'Language models are realistic tabular data generators'. In: *arXiv preprint arXiv:2210.06280* (2022) (cited on page 211).

[156] Jian-Qiao Zhu and Thomas L Griffiths. 'Eliciting the priors of large language models using iterated in-context learning'. In: *arXiv preprint arXiv:2406.01860* (2024) (cited on page 211).

[157] Marcel Binz and Eric Schulz. 'Using cognitive psychology to understand GPT-3'. In: *Proceedings of the National Academy of Sciences* 120.6 (2023), e2218523120 (cited on page 211).

[158] Edgar A. Duéñez-Guzmán et al. 'A Social Path to Human-like Artificial Intelligence'. In: *Nature Machine Intelligence* 5.11 (Nov. 2023), pp. 1181–1188. DOI: 10.1038/s42256-023-00754-x. (Visited on 12/03/2023) (cited on page 211).

[159] Marcel Binz et al. 'Heuristics from bounded meta-learned inference.' In: *Psychological review* 129.5 (2022), p. 1042 (cited on page 211).

[160] Berndt Brehmer. 'Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks'. In: *Organizational Behavior and Human Performance* 11.1 (1974), pp. 1–27 (cited on page 212).

[161] Berndt Brehmer. 'Single-cue probability learning as a function of the sign and magnitude of the correlation between cue and criterion'. In: *Organizational Behavior and Human Performance* 9.3 (1973), pp. 377–395 (cited on page 212).

[162] Berndt Brehmer, Jan Kuylenstierna, and Jan-Erik Liljergren. 'Effects of function form and cue validity on the subjects' hypotheses in probabilistic inference tasks'. In: *Organizational Behavior and Human Performance* 11.3 (1974), pp. 338–354 (cited on page 212).

[163] Edward L DeLosh, Jerome R Busemeyer, and Mark A McDaniel. 'Extrapolation: the sine qua non for abstraction in function learning.' In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23.4 (1997), p. 968 (cited on page 212).

[164] Eunhee Byun. 'Interaction between prior knowledge and type of nonlinear relationship on function learning'. PhD thesis. Purdue University, 1995 (cited on page 212).

[165] J Douglas Carroll. 'Functional learning: The learning of continuous functional mappings relating stimulus and response continua'. In: *ETS Research Bulletin Series* 1963.2 (1963), pp. i–144 (cited on page 212).

[166] Mark A Mcdaniel and Jerome R Busemeyer. 'The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models'. In: *Psychonomic bulletin & review* 12.1 (2005), pp. 24–42 (cited on page 212).

[167] Peter J Kwantes and Andrew Neal. 'Why people underestimate y when extrapolating in linear functions.' In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32.5 (2006), p. 1019 (cited on page 212).

[168] Mark K Johansen and Thomas J Palmeri. 'Are there representational shifts during category learning?' en. In: *Cogn. Psychol.* 45.4 (Dec. 2002), pp. 482–553. DOI: 10.1016/s0010-0285(02)00505-4 (cited on page 212).

[169] Akshay K Jagadish et al. 'Human-like Category Learning by Injecting Ecological Priors from Large Language Models into Neural Networks'. In: *Forty-first International Conference on Machine Learning*. 2024 (cited on page 214).

[170] Johannes A Schubert et al. 'In-context learning agents are asymmetric belief updaters'. In: *Forty-first International Conference on Machine Learning*. 2024 (cited on page 214).

[171] Akshay Kumar Jagadish et al. *Zero-Shot Compositional Reinforcement Learning in Humans*. July 2023. DOI: 10.31234/osf.io/ymve5. (Visited on 07/26/2023) (cited on page 214).

[172] William Vallet and Virginie van Wassenhove. 'Can cognitive neuroscience solve the lab-dilemma by going wild?' In: *Neuroscience & Biobehavioral Reviews* (2023), p. 105463 (cited on page 214).

[173] Tal Yarkoni. 'The generalizability crisis'. In: *Behavioral and Brain Sciences* 45 (2022), e1 (cited on page 214).

[174] Sang Michael Xie et al. 'An explanation of in-context learning as implicit bayesian inference'. In: *arXiv preprint arXiv:2111.02080* (2021) (cited on page 214).

[175] Marcel Binz et al. 'Centaur: a foundation model of human cognition'. In: *arXiv preprint arXiv:2410.20268* (2024) (cited on pages 214, 216, 219).

[176] Murray Shanahan et al. 'Artificial intelligence and the common sense of animals'. In: *Trends in Cognitive Sciences* 24.11 (2020), pp. 862–872 (cited on page 215).

[177] Kelsey Allen et al. 'Using games to understand the mind'. In: *Nature Human Behaviour* (2024), pp. 1–9 (cited on page 215).

[178] Adaptive Agent Team et al. *Human-Timescale Adaptation in an Open-Ended Task Space*. Jan. 2023. (Visited on 01/30/2023) (cited on page 215).

[179] Mojang Studios Markus Persson. *Minecraft*. First released in 2011, current version by Mojang Studios, a subsidiary of Microsoft. 2011. URL: https://minecraft.net (cited on page 215).

[180] Maximilian Beck et al. 'xlstm: Extended long short-term memory'. In: *arXiv preprint arXiv:2405.04517* (2024) (cited on page 215).

[181] Kazuki Irie, Róbert Csordás, and Jürgen Schmidhuber. 'Metalearning continual learning algorithms'. In: *arXiv [cs.LG]* (Nov. 2023) (cited on pages 215, 216).

[182] Jürgen Schmidhuber. 'A 'self-referential'weight matrix'. In: *ICANN'93: Proceedings of the International Conference on Artificial Neural Networks Amsterdam, The Netherlands 13–16 September 1993 3*. Springer. 1993, pp. 446–450 (cited on page 215).

[183] Jane X. Wang et al. *Learning to Reinforcement Learn*. Jan. 2017. (Visited on 10/05/2022) (cited on pages 215, 216).

[184] Kristopher T Jensen, Guillaume Hennequin, and Marcelo G Mattar. 'A recurrent network model of planning explains hippocampal replay and human behavior'. In: *Nature neuroscience* 27.7 (2024), pp. 1340–1348 (cited on page 216).

[185] Sreejan Kumar et al. 'Using natural language and program abstractions to instill human inductive biases in machines'. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 167–180 (cited on page 216).

[186] Inc. Meta Platforms. *Llama 3.1 70B Model*. 2024. URL: https://github.com/meta-llama/llama3 (cited on page 216).

[187] Jeffrey S. Bowers et al. *Centaur: A model without a theory*. Preprint. 2025. DOI: 10.31234/osf.io/v9w37.v2. URL: https://doi.org/10.31234/osf.io/v9w37.v2 (cited on page 216).

[188] Pacific Laboratory for Artificial Intelligence. *plaicraft.ai*. en. https://plai.cs.ubc.ca/2023/09/27/plaicraft/. Accessed: 2024-9-10. Sept. 2023 (cited on page 216).

[189] Roger Ratcliff. 'Connectionist models of recognition memory: constraints imposed by learning and forgetting functions.' In: *Psychological review* 97.2 (1990), p. 285 (cited on page 216).

[190] Michael McCloskey and Neal J Cohen. 'Catastrophic interference in connectionist networks: The sequential learning problem'. In: *Psychology of learning and motivation*. Vol. 24. Elsevier, 1989, pp. 109–165 (cited on page 216).

[191] Samuel Ritter et al. 'Been there, done that: Meta-learning with episodic recall'. In: *International conference on machine learning*. PMLR. 2018, pp. 4354–4363 (cited on page 216).

[192] Kazuki Irie and Brenden M Lake. 'Neural networks that overcome classic challenges through practice'. In: *arXiv preprint arXiv:2410.10596* (2024) (cited on page 216).

[193] Rishav Mukherji et al. 'Activity sparsity complements weight sparsity for efficient RNN inference'. In: *arXiv preprint arXiv:2311.07625* (2023) (cited on page 217).

[194] Guillaume Bellec et al. 'Long short-term memory and learning-to-learn in networks of spiking neurons'. In: *Advances in neural information processing systems* 31 (2018) (cited on page 217).

[195] Robert Tjarko Lange and Henning Sprekeler. 'Learning Not to Learn: Nature versus Nurture In Silico'. en. In: *AAAI* 36.7 (June 2022), pp. 7290–7299. DOI: 10.1609/aaai.v36i7.20691 (cited on page 217).

[196] David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., 1982 (cited on page 217).

[197] Ryoma Hattori et al. 'Meta-reinforcement learning via orbitofrontal cortex'. In: *Nature Neuroscience* 26.12 (2023), pp. 2182–2191 (cited on page 217).

[198] Hoagy Cunningham et al. 'Sparse autoencoders find highly interpretable features in language models'. In: *arXiv preprint arXiv:2309.08600* (2023) (cited on page 217).

[199] Trenton Bricken et al. 'Towards monosemanticity: Decomposing language models with dictionary learning'. In: *Transformer Circuits Thread* 2 (2023) (cited on page 217).

[200] Can Demircan et al. 'Sparse autoencoders reveal temporal difference learning in large language models'. In: *arXiv preprint arXiv:2410.01280* (2024) (cited on page 217).

[201] Thomas Naselaris et al. 'Encoding and decoding in fMRI'. In: *Neuroimage* 56.2 (2011), pp. 400–410 (cited on page 217).

[202] Justin N Wood et al. 'Reverse engineering the origins of visual intelligence'. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 42. 2020 (cited on page 218).

[203] Nora S Newcombe. 'Cognitive development: changing views of cognitive change'. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 4.5 (2013), pp. 479–491 (cited on page 218).

[204] Dima Amso, Carmel Salhi, and David Badre. 'The relationship between cognitive enrichment and cognitive control: A systematic investigation of environmental influences on development through socioeconomic status'. In: *Developmental psychobiology* 61.2 (2019), pp. 159–178 (cited on page 218).

[205] Adriana Galván. 'Neural plasticity of development and learning'. In: *Human brain mapping* 31.6 (2010), pp. 879–890 (cited on page 218).

[206] Jamie L Hanson et al. 'Impact of early life stress on reward circuit function and regulation'. In: *Frontiers in Psychiatry* 12 (2021), p. 744690 (cited on page 219).

[207] Rasmus M Birn, Barbara J Roeber, and Seth D Pollak. 'Early childhood stress exposure, reward pathways, and adult decision making'. In: *Proceedings of the National Academy of Sciences* 114.51 (2017), pp. 13549–13554 (cited on page 219).

[208] A David Redish and Joshua A Gordon. *Computational psychiatry: New perspectives on mental illness*. Vol. 20. 2022 (cited on page 219).

[209] Quentin JM Huys, Tiago V Maia, and Michael J Frank. 'Computational psychiatry as a bridge from neuroscience to clinical applications'. In: *Nature neuroscience* 19.3 (2016), pp. 404–413 (cited on page 219).

[210] Anita Woolfolk. *Educational psychology*. Pearson, 2016 (cited on page 219).

[211] William Edwin Segall and Anna Victoria Wilson. *Introduction to education: Teaching in a diverse society*. Rowman & Littlefield, 2004 (cited on page 219).

[212] EsraEret1 Tuba Gokmenoglu CennetEngin-Demir. 'A review of research on educational theories and approaches affecting students achievement: 1990-2011'. In: *Elementary Education Online* 12.3 (2013), pp. 687–700 (cited on page 219).

[213] Trudi Cooper. 'Curriculum Renewal: Barriers to Successful Curriculum Change and Suggestions for Improvement.' In: *Journal of Education and Training Studies* 5.11 (2017), pp. 115–128 (cited on page 219).

[214] Kinza Aslam et al. 'Curriculum implementation challenges: Development and validation of an integrated curriculum implementation challenges tool'. In: *Pakistan Journal of Medical Sciences* 40.1Part-I (2024), p. 89 (cited on page 219).

[215] Jason Wei et al. 'Emergent abilities of large language models'. In: *arXiv preprint arXiv:2206.07682* (2022) (cited on page 219).

[216] Marcel Binz et al. 'Meta-learning: Data, architecture, and both.' In: *Behavioral & Brain Sciences* 47 (2024) (cited on page 219).

[217] OpenAI. *Learning to reason with LLMs*. en. https://openai.com/index/learning-to-reason-with-llms/. Accessed: 2025-4-15. 2024 (cited on page 219).

[218] Milena* Rmus et al. 'Generating Computational Cognitive Models using Large Language Models'. In: *arXiv preprint arXiv:2502.00879* (2025). *These authors contributed equally as joint first authors (cited on page 219).

[219] Jacques Pesnot Lerousseau and Christopher Summerfield. 'Quo Vadis, Planning?' In: *OSF* (2024) (cited on page 220).