META-LEARNING OF COMPOSITIONAL TASK DISTRIBUTIONS IN HUMANS AND MACHINES

Anonymous authors
Paper under double-blind review

ABSTRACT

Modern machine learning systems struggle with sample efficiency and are usually trained with enormous amounts of data for each task. This is in sharp contrast with humans, who often learn with very little data. In recent years, meta-learning, in which one trains on a family of tasks (i.e. a task distribution), has emerged as an approach to improving the sample complexity of machine learning systems and to closing the gap between human and machine learning. However, in this paper, we argue that current meta-learning approaches still differ significantly from human learning. We argue that humans learn over tasks by constructing compositional generative models and using these to generalize, whereas current meta-learning methods are biased toward the use of simpler statistical patterns. To highlight this difference, we construct a new meta-reinforcement learning task with a compositional task distribution. We also introduce a novel approach to constructing a "null task distribution" with the same statistical complexity as the compositional distribution but without explicit compositionality. We train a standard meta-learning agent, a recurrent network trained with model-free reinforcement learning, and compare it with human performance across the two task distributions. We find that humans do better in the compositional task distribution whereas the agent does better in the non-compositional null task distribution – despite comparable statistical complexity. This work highlights a particular difference between human learning and current meta-learning models, introduces a task that displays this difference, and paves the way for future work on human-like meta-learning.

1 INTRODUCTION

Modern machine learning models are typically trained using an enormous number of examples on a specific task. This is in sharp contrast to humans, who instead learn efficiently from very few examples (Lake et al., 2015). There are two principles underlying humans' ability to do this. First, the ability to learn and represent information that transfers across tasks. Humans can "learn-to-learn", allowing them to improve their learning strategies over time and learn new tasks quickly from few examples (Thrun & Pratt, 1998; Lake et al., 2017). These acquired strategies or inductive biases constrain the space of possibilities, making learning more data efficient (Lake et al., 2015). Second, humans represent this learned information compositionally. These abstract representations allow for generalization far outside the training regime to more complex environments via recursive combination of simpler building blocks (Kemp & Tenenbaum, 2008; Schulz et al., 2017).

Inspired by the first principle above (Wang et al., 2016; Botvinick et al., 2019), as well as by the broader engineering problem of improving sample complexity (Wang et al., 2020), researchers have renewed investigations into the idea of meta-learning. Under this approach, a model is trained on a family of tasks to learn an inductive bias that helps it learn a new task from the same task distribution using fewer examples (Hospedales et al., 2020). In this work, we directly examine the second principle outlined above: whether the inductive biases learned via meta-learning are compositional.

Previous work demonstrates that some compositionality can be meta-learned (Lake, 2019). Our work is distinct in three main ways. First, we explore compositionality in an explicit generative grammar, with a reinforcement learning task. Second, we provide a rigorous comparison between human and meta-learning agent behavior on our task. Third, our focus is on whether the agent meta-

learns a strategy or inductive bias that represents abstract compositional rules or if it instead exploits statistical patterns that are a consequence of those rules. This third point is especially crucial – we argue that simply doing well on a compositional task distribution does not necessarily indicate meta-learning compositional structure, and provide a direct way to disentangle compositionality from statistical patterns.

We hypothesise that while models of meta-learning do pick up relevant information from the training task distribution, they are biased toward acquiring statistical patterns rather than compositional rules. This is in contrast to humans, who have an inductive bias toward acquiring and representing compositional structure.[1] Our methodological contribution in this work is develop novel tasks to examine if agents meta-learn compositionality, while directly controlling for statistical complexity. We also present several new empirical results that directly compare the performance of a meta-learned agent with human performance on two tasks: one that has compositional structure, and one that is of equal statistical complexity but does not have compositional structure. We show through three different analyses that humans have a bias toward compositional representations and easily pick up this structure within a short task. In contrast, standard meta-learning agents consistently do not acquire this inductive bias, despite extensive training on compositional task distributions – they instead pick up statistical patterns that are a consequence of this compositional structure.

This finding highlights that, although meta-learning is a very powerful approach for endowing artificial systems with useful inductive biases (McCoy et al., 2020; Dasgupta et al., 2019; Lake, 2019; Griffiths et al., 2019), there still exist multiple ways to represent this abstract knowledge. Just as every vanilla learning algorithm has its own inductive bias in learning a single task, meta-learners themselves have inductive biases (which we term *meta-inductive bias*) influencing what inductive biases they learn from their cross-task experience. In this paper, we show that standard meta-learners do not learn a compositional bias despite direct training on compositional task distributions. Since a large swath of real-world tasks contain compositional structure, this bias is objectively valuable – not simply a way to be more human-like. This work highlights that this valuable inductive bias cannot easily be learned by standard meta-learners.

2 Embedding compositionality in a task distribution

In this work, we define a broad family of task distributions that contain abstract compositional structure. Previous work on such datasets (Lake & Baroni, 2018; Johnson et al., 2017) focuses primarily on language. Here we instead directly consider the domain of structure learning. This is a fundamental tenet of human cognition and has been linked to how humans learn quickly in novel environments (Tenenbaum et al., 2011; Mark et al., 2020). Structure learning is required in a vast range of domains: from planning (understanding an interrelated sequence of steps for cooking), category learning (the hierarchical organization of biological species), to social inference (understanding a chain of command at the workplace, or social cliques in a high school). A task distribution based on structure learning can therefore be embedded into several domains relevant for machine learning.

Kemp & Tenenbaum (2008) provide a model for how people infer such structure. They present a probabilistic context-free graph grammar that produces a space of possible structures, over which humans do inference. A grammar consists of a start symbol S, terminal and non-terminal symbols Σ and V, as well as a set of production rules R. Different structural forms arise from recursively applying these production rules. This framework allows us to specify abstract structures (via the grammar) and to produce various instantiations of this abstract structure (via the noisy generation process), naturally producing different families of task distributions.

We consider three structures: chains, trees, and loops. These exist in the real world across multiple domains. Chains describe objects on a one-dimensional spectrum, like people on the left-right political spectrum. Trees describe objects organized in hierarchies, like evolutionary trees. Loops describe cycles, like the four seasons. Here we embed these structures into a grid-based task.

Exploration on a grid is an extensively studied problem in machine learning, particularly in reinforcement learning. Further, it is also a task that is easy for humans to perform on online crowd-

[1] We do not address how humans came to have this inductive bias. In this work, we study whether meta-learning agents can learn this bias given training on compositional task distributions.

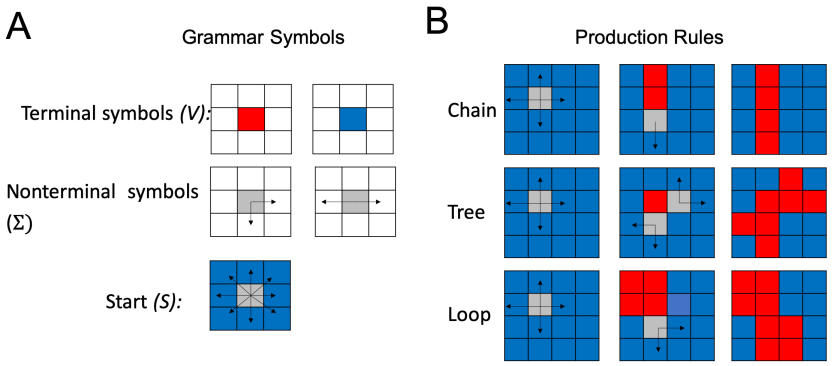

Figure 1: **Generative Grammar** (A) Grammar symbols and (B) production rules. A board is formed by beginning with the start symbol and recursively applying production rules until only terminal symbols (red and blue tiles) are left. Each production rule either adds a non-terminal symbol (from first column to second) or a terminal symbol (from second column to third) with 0.5 probability.

sourcing platforms – but not trivially so. This allows us to directly compare human and machine performance on the same task. Fig. 1 displays the symbols of the grammar we use and the production rules that give rise to grids of different structural forms.

2.1 A TASK TO TEST STRUCTURE LEARNING

Here we describe the specific task built atop this embedding of structural forms. We use a tile revealing task on the grid. Humans as well as agents are shown a 7x7 grid of tiles, which are initially white except for one red tile. This red tile is the initial start tile in the grid's generative process (see Fig. 1). Clicking white tiles reveal them to be either red or blue. The episode finishes when the agent reveals all the red tiles. There is a reward for each red tile revealed, and a penalty for every blue tile revealed. The goal therefore is to reveal all the red tiles while revealing as few blue tiles as possible. The particular configuration of the red tiles defines a single task. The distribution of tasks for meta-learning is defined by the grammar from which these structures are sampled. Here, we randomly sampled from a uniform mixture of chains, trees, and loops as defined in Fig. 1.

2.2 A STATISTICALLY EQUIVALENT NON-COMPOSITIONAL TASK

Previous approaches to testing compositionality in machine-learned representations (Lake & Baroni, 2018; Dasgupta et al., 2018) have relied on examining average performance on held-out examples from compositionally structured task distributions. However, we argue that this often confounds whether a system has truly acquired compositional structure or whether it is relying on statistical structure that comes about as a consequence of compositional rules.

To directly examine whether compositionality is a factor in how humans and meta-learning agents perform this task, we need a control task distribution that is similar in statistical complexity but is not explicitly compositional. To this end, we trained a fully connected neural network (3 layers, 49 units each) to learn the conditional distribution of each tile given the all other tiles on the compositional boards. Note that these conditional distributions contain all the relevant statistical information about the boards. We do this by training on an objective inspired by masked language models like BERT (Devlin et al., 2018). The network was given a compositional board with a random tile masked out and trained to reproduce the entire board including the randomly masked tile. The loss was binary cross entropy between the predicted and actual masked tiles. The network was trained on all possible compositional boards for 10^4 epochs, training accuracy was $\sim$99%.

We then sampled boards from these conditionals with Gibbs sampling. We started with a grid where each tile is randomly set to red or blue with probability 0.5. We then masked out a tile and ran the grid through the network to get the conditional probability of the tile being red given the other tiles, turning the tile red with that probability. We repeated this by masking each tile in the 7x7 grid (in a random order) to complete a single Gibbs sweep, and repeat this whole Gibbs sweep 20 times to

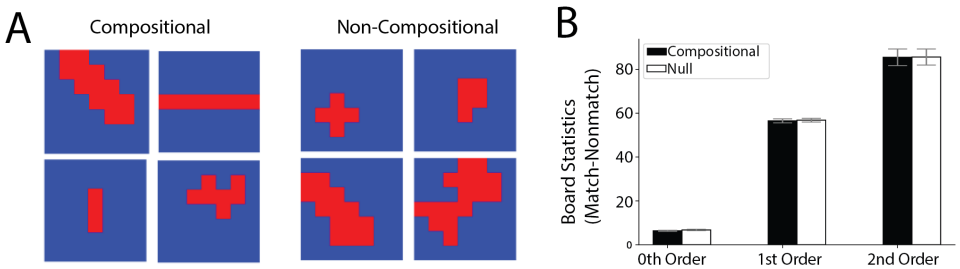

Figure 2: **Comparing compositional and null task distributions** (A) Example compositional and null distribution boards. Compositional boards are distinctly either chain, trees, or loops while null boards have similar statistical properties but don't necessarily obey the recursive rules used to generate compositional boards. (B) Ising statistics across the two task distributions. Error bars are 95% non-parametric bootstrap confidence intervals across different boards of the respective distribution.

generate a single sample. We refer to the distribution of boards generated this way as the null task distribution. Fig. 2 shows example compositional and null distribution grids.

While the statistical structure looks similar, the non-compositional null boards shown could not have been generated by the grammar in Fig. 1. The conditional distributions for the two distributions are similar by design, we further quantify statistical similarity using Ising statistics (Zhang, 2007). We compared the 0th order, 1st order, and 2nd order effects defined as follows. The 0th order statistic corresponds to the number of red minus number of blue tiles. The 1st order statistic counts the number of agreeing neighbours (vertically or horizontally adjacent) minus the disagreeing ones, where agreeing means being of the same color. The 2nd order statistic is the number of triples (tile+its neighbor+its neighbor's neighbor) that agree, minus those that don't. Fig. 2b shows that the two distributions are not significantly different in terms of the Ising statistics measured ($p > 0.05$ for all three orders).

3 Experiments

We analyze and compare the performance of standard meta-learners and human learning on our tile-revealing task. We test them on boards that are sampled from the generative grammar and contain explicit compositional structure, as well as on boards that are matched for statistical complexity, but are sampled from a null distribution that does not contain explicit compositional structure. Comparing performance across these two task distributions allows us to pinpoint the role of compositionality as distinct from the statistical patterns that arise as a downstream consequence of compositional rules.

3.1 Methods

Meta-Reinforcement Learning Agent Following previous work in meta-reinforcement learning (Wang et al., 2016; Duan et al., 2016) we use an LSTM meta-learner that takes the full board as input, passes it through 2 fully connected layers (49 units each) and feeds that, along with the previous action and reward, to 120 LSTM units. It is trained with a linear learning rate schedule and 0.9 discount. The reward function was: +1 for revealing red tiles, -1 for blue tiles, +10 for the last red tile, and -2 for choosing an already revealed tile. The agent was trained using Advantage Actor Critic (A2C) (Stable baselines package Hill et al., 2018). The agent was trained for 10^6 episodes. We performed a hyperparamater sweep (value function loss coefficient, entropy loss coefficient, learning rate) using a held-out validation set for evaluation (see Appendix). The selected model's performance was evaluated on held-out test grids. We trained different agents in the same way on the compositional and null task distributions, with separate hyperparameter sweeps for each.

Human Experiment We crowdsourced human performance on our task using Prolific (www.prolific.co) for a compensation of $1.50. Participants were shown the 7x7 grid on their web

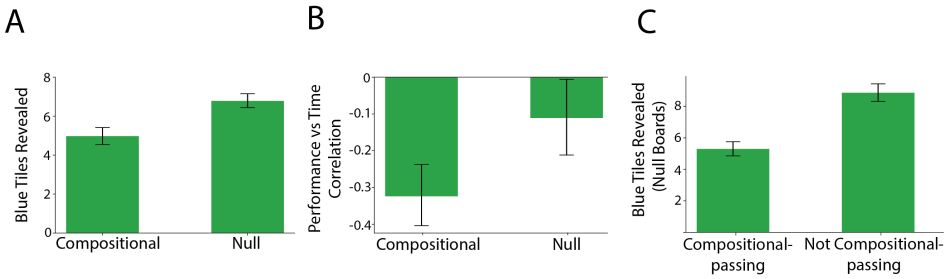

Figure 3: **Human performance.** (A) Humans perform better (i.e. less blue tiles) in the compositional vs null task distribution ($p<0.0001$). (B) Human performance improves over the course of the experiment (indicated by negative correlation between trial number and number of blue tiles revealed), with significantly greater improvement for compositional distribution ($p=0.0006$). (C) Some null distribution boards can pass as compositional – humans perform significantly better on these than on other boards in the null distribution ($p<0.0001$).

browser and used mouse-clicks to reveal tiles. Each participant was randomly assigned to the compositional or null task distribution, 25 participants in each. Each participant was directly evaluated on the test set of grids for the models (24 grids from their assigned task distribution in randomized order). This is a key difference between the human and agent tasks – humans did not receive training on a task distribution. While we are examining whether agents can *meta-learn* an inductive bias toward compositionality (by training on compositional task distributions), we assume that humans already have this bias from pre-experimental experience. Since participants had to reveal all red tiles to move on to the next grid, they were implicitly incentivized to be efficient (clicking as few blue tiles as possible) in order to finish the task quickly. We found that this was adequate to get good performance. A reward structure similar to that given to agents was displayed as the number of points accrued, but did not translate to monetary reward.

Evaluation Unless specified otherwise, performance is evaluated as the number of blue tiles revealed before all red tiles are revealed (lower is better). All error bars are 95% non-parametric bootstrap confidence intervals calculated across agents / participants. Non-overlapping confidence intervals will have a significant difference, but we also include non-parametric bootstrapped p-values for differences across different samples (e.g. human vs agent).

3.2 Results

In this section, we first describe human behavior on this novel task. We demonstrate that humans have a clear bias toward compositional distributions, without extensive training and even while directly controlling for statistical complexity. We then compare human performance with that of a meta-learning agent—which has had extensive training on this task, and therefore has had the chance to learn the inductive biases relevant to this task distribution. We find significant qualitative and quantitative differences in behavior, and examine the role of meta-inductive bias – i.e. what kinds of abstract cross-task structure do meta-learners prefer to represent? In particular, we consider compositional and spatial inductive biases. Finally, we demonstrate the effect of an architectural change (adding convolutions) in the meta-learner that makes it easier for it to discover spatial structure. We demonstrate that, while this helps agent performance overall, it further highlights the divergence between human and agent behavior along the dimension of compositionality.

Human performance: We found that participants do better on the compositional task distribution than the null task distribution (see Fig. 3a). Despite not having been trained on this task beforehand, human participants do fairly well on this task from the get go indicating that humans might have some of the relevant inductive biases from pre-experimental experience. To test if there is still learning within the experiment, we correlated trial number with the number of blue tiles revealed (Fig. 3B), and found improvement across both conditions but significantly greater improvement for the compositional distribution. Finally, we investigate performance on the null task distribution more closely. There is some overlap between the null and compositional distributions, because some of the

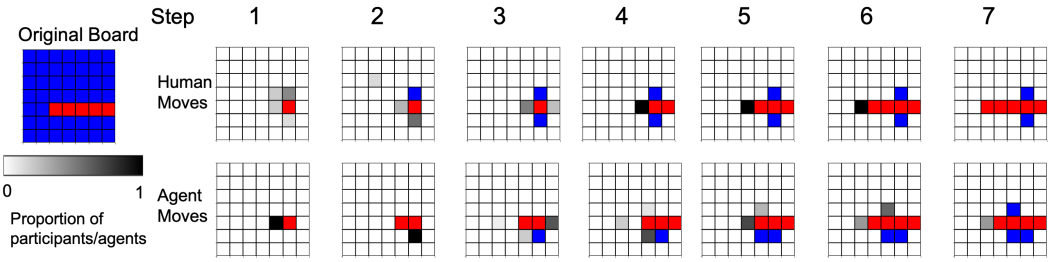

Figure 4: **Human and agent policies on the task.** Red/blue indicate already revealed tiles while grayscale indicate what proportion of humans or agents (conditioned on the moves so far) revealed that tile in the next step. In this chain example, once humans figure out that the board is a chain structural form (step 5), they get perfect performance by choosing tiles along the chain's production direction, while agents still choose other blue tiles.

generated null boards could have been generated by the production rules of the generative grammar. We split the null test set by whether or not the board is 'compositional-passing' and compare human performance across these. We find that humans do significantly better on boards that could have have been generated by the compositional production rules (Fig. 3c). This further indicates an inductive bias that aligns with the compositional distribution over the null distribution.

Comparing human and agent performance: First, we note that the meta-learners perform relatively well on this task (Fig. 5), indicating that they have learned some generalizable information from the distribution of tasks. Since the test set has held-out board of a compositional grammar, traditional approaches to evaluating compositionality might declare victory. However, in this paper, our goal is to decouple whether the agent truly learns an inductive bias toward compositional rules like humans, or if it instead learns statistical patterns that are a consequence of compositional rules.

We start with an example that highlights the difference between human and agent policies on this task (Fig. 4). In this chain example, once humans figure out that the board is a chain structural form, they never deviate from the chain's production direction while agents do. This indicates that humans are learning the generative rules of the chain form and using these rule to determine their actions, while the agent is using simpler statistical patterns that do not have strict rules.

We now consider various ways to quantify this difference. First, we see that humans do better overall on both the null and compositional distributions (Fig. 5; $p<0.0001$ for both task distributions). This is despite, unlike the agents, no direct experience with this task. This indicates that humans have useful inductive biases from pre-experimental experience that are valuable in this task (Dubey et al., 2018). We discuss the role of these inductive biases in the following sections. The meta-learner has had extensive experience with each task distribution, and had the chance to pick up the inductive biases relevant for this task. Persistent differences in performance indicate that standard meta-learners differ from humans in the kinds of inductive biases they learn (i.e. in their meta-inductive biases).

Bias toward compositionality. First, we note that humans perform better on the compositional versus the null distribution (Fig. 5a), whereas the agent does better on the null task distribution than on the compositional tasks. This reflects a significant difference between their performance. We hypothesized that humans perform well on the compositional task distribution by first inferring what kind of structure they are in, and then following the production rules for that structure. Since such structure does not reliably exist in the null distribution, they cannot learn, infer, and use it. Further, we hypothesized that the agents learn statistical patterns instead.

Fig. 4 supports this intuition but here we look to quantify it further. If a system uses a compositional generative representation, we would expect success rate (rate of choosing red tiles) to be low in the beginning of a trial while figuring out what the structure underlying this trial is. Conversely, we would expect a higher success rate towards the end while following inferred production rules to reveal red tiles. To test this hypothesis, we split human and agent behavior in each trial into the first and last half, and examine success rate in each Fig. (5b and c). For the compositional

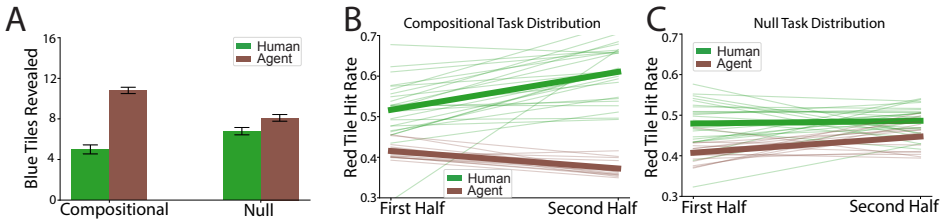

Figure 5: **Comparing human and agent performance** (A) Humans do better at the compositional task than the null ($p<0.0001$), while agents do better at null ($p<0.0001$). (B) Humans have a higher success rate revealing red tiles in the second half of a trial for the compositional task ($p<0.0001$), agents do not. Transparent line represents individual human/agent average over trials, thick lines represent average over humans/agents. (C) Humans do not improve their success rates during a trial in the null task while agents do ($p=0.0014$).

distribution, we find that humans have a higher success rate in the second half, providing support for our hypothesis. In contrast, we find that agent success rate does not increase over a trial, and in fact decreases. We also find that humans do not show increasing success rate in the null task distribution while agents do, providing further evidence for our hypothesis.

Bias toward spatial proximity. We note that humans outperform the agent even in the null task distribution, despite extensive training for the agent. One possibility is that good human performance in the null task is explained by their performance on the compositional-passing examples in the null task distribution (Fig. 3c). However, another possibility is that humans come into the task with strong inductive biases about spatial proximity.[2] While the starting tile for the grammar can be randomly chosen, the production rules operate over nearest-neighbour adjacencies. A system that has a bias toward local spatial structure might therefore find this task easier.

We test this intuition by comparing performance to a heuristic that uses only local spatial information. This heuristic selects uniformly from the (unrevealed) nearest neighbors of a randomly selected red tile. We evaluated this heuristic 1,000 times on each test board and formed a z-score statistic by subtracting the mean heuristic performance from the human's/agent's performance for each board divided by the standard deviation of the heuristic's performance. We find that humans do better than the neighbor heuristic (Fig.6a), while the agent does not.

We can give a neural network a bias toward spatial proximity using convolutions (LeCun et al., 1989). To test if this helps the agent, we replaced the agent's first fully connected layer with a convolutional layer. We find that this agent outperforms humans on the null task distribution(Fig.6b). We also find that it outperforms the spatial heuristic described above (Fig.6c). Note that this strictly reduces the expressivity of the model, and any improvements are due to the right inductive bias.

However, humans still perform better than the agent in the compositional task distribution. This result provides a double dissociation between statistical learning and compositional inductive bias. It shows that the gap between humans and agents on the compositional task is not due to artificial meta-learners being overall worse learners than humans – the convolutional meta-learner actually outperforms humans on the null distribution of equal statistical complexity. This provides further evidence that the inductive bias toward and representing compositional structure is what gives humans a competitive advantage over the agent on these compositional task distributions, and that meta-learners do not learn it despite access to compositional training distributions.

4 Discussion

Compositionality is a central tenet of human intelligence, allowing for the infinite use of finite means (Fodor et al., 1988; Lake et al., 2017). An inductive bias toward these widely generalizable

[2]Spatial structure is shared by both distributions (Fig. 2) and can't explain why humans are better at compositional while agents are better at null. We investigate if it can explain why humans perform better overall.

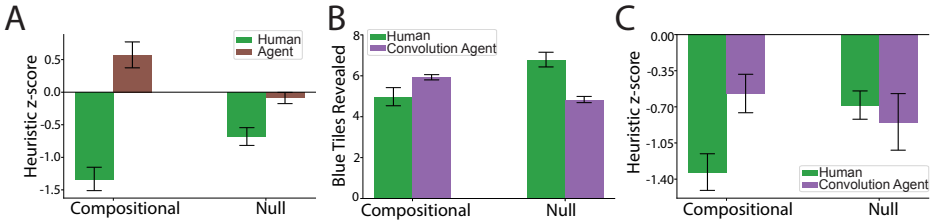

Figure 6: **Role of spatial structure in task performance.** (A) Humans outperform the neighbor heuristic (negative z-score), the agent performs worse. (B) Humans outperform the convolution agent in the compositional distribution ($p<0.0001$) and the convolution agent outperforms humans in the null distribution ($p<0.0001$). (C) The convolution agent outperforms neighbor heuristic.

representations would be of great value to machine learning systems. Recent developments in meta-learning hold promise as an approach to endowing systems with such useful inductive biases. This work makes several methodological and scientific contributions to provide a rigorous way to test for a compositional inductive bias. We show that popular meta-learning approaches, in sharp contrast to humans, struggle with compositionality and prefer statistical patterns.

Our first contribution is to develop compositionally structured task distributions for meta-learning using explicit generative grammars (Kemp & Tenenbaum, 2008). Previous work on generating compositional datasets has focused on language. We argue that using explicit generative grammars has the dual advantage of being generalizable to a variety of structure, as well as being easy to embed in multiple domains relevant to machine learning. In this work, we embed this structure into a grid-based task. Grid-based tasks are commonly studied in reinforcement-learning, are easy for humans to perform on online platforms, and behavior on this task is easy to visualize and interpret. This provides fertile ground for direct comparisons between human and machine behavior, as we demonstrate in our experiments. Previous work on meta-learning compositionality uses performance on a compositional task distribution as an indicator for meta-learning this inductive bias (Lake, 2019). However, it is possible for meta-learning systems to perform well using statistical patterns instead. Our second methodological contribution is to create distributions with comparable statistical complexity to compositional distributions, that are not explicitly compositional. This control distribution allows us to disentangle statistical pattern matching from rule-based compositionality. Our method uses a neural network to learn conditional distributions and generate Gibbs samples. This approach is similar to masked language modelling (Devlin et al., 2018), and our findings—that this procedure generates statistically similar but non-compositional distributions, that are in fact *easier* for downstream networks to learn than the true compositional distribution—are also relevant to understanding the representations learned by these systems more broadly (Rogers et al., 2020).

In our experiments, we first show that humans have a bias toward compositional structure, while directly controlling for statistical complexity. This generalizes findings in the space of function learning (Schulz et al., 2017) to grid-based reinforcement learning tasks. Further, we find that agents (recurrent network trained with model-free reinforcement learning, following Wang et al., 2016; Duan et al., 2016) find the non-compositional distribution easier to learn than the compositional one. This is in direct contrast with human behavior, indicating that agents do not acquire this inductive bias through meta-learning. A followup experiment with a convolutional agent directly dissociates statistical learning prowess from a compositional inductive bias, and highlights compositionality as the key difference between humans and agents in this task.

While previous work has demonstrated that some aspects of compositionality can be meta-learned with training on extensive data (Lake, 2019), our focus here is to demonstrate that compositional structure remains *difficult* for these systems – and that they prefer other statistical features when possible. In other words, they do not have a meta-inductive bias toward learning compositionality as an inductive bias. This highlights the importance of endowing artificial systems with this bias. Meta-learning with graph neural networks (Battaglia et al., 2018) or neurosymbolic approaches (Ellis et al., 2020) is promising. An exciting direction for future work is to examine a range of approaches to learning compositional representations with the tools we set forth in this paper, and using the resulting insights to move toward closing the gap between human and machine intelligence.

REFERENCES

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Matthew Botvinick, Sam Ritter, Jane X Wang, Zeb Kurth-Nelson, Charles Blundell, and Demis Hassabis. Reinforcement learning, fast and slow. *Trends in cognitive sciences*, 23(5):408–422, 2019.

Ishita Dasgupta, Demi Guo, Andreas Stuhlmüller, Samuel J Gershman, and Noah D Goodman. Evaluating compositionality in sentence embeddings. *arXiv preprint arXiv:1802.04302*, 2018.

Ishita Dasgupta, Jane Wang, Silvia Chiappa, Jovana Mitrovic, Pedro Ortega, David Raposo, Edward Hughes, Peter Battaglia, Matthew Botvinick, and Zeb Kurth-Nelson. Causal reasoning from meta-reinforcement learning. *arXiv preprint arXiv:1901.08162*, 2019.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Yan Duan, John Schulman, Xi Chen, Peter L Bartlett, Ilya Sutskever, and Pieter Abbeel. Rl˜2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*, 2016.

Rachit Dubey, Pulkit Agrawal, Deepak Pathak, Thomas L Griffiths, and Alexei A Efros. Investigating human priors for playing video games. *arXiv preprint arXiv:1802.10217*, 2018.

Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sable-Meyer, Luc Cary, Lucas Morales, Luke Hewitt, Armando Solar-Lezama, and Joshua B Tenenbaum. Dreamcoder: Growing generalizable, interpretable knowledge with wake-sleep bayesian program learning. *arXiv preprint arXiv:2006.08381*, 2020.

Jerry A Fodor, Zenon W Pylyshyn, et al. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3–71, 1988.

Thomas L Griffiths, Frederick Callaway, Michael B Chang, Erin Grant, Paul M Krueger, and Falk Lieder. Doing more with less: meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, 29:24–30, 2019.

Ashley Hill, Antonin Raffin, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, Rene Traore, Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, and Yuhuai Wu. Stable baselines. `https://github.com/hill-a/stable-baselines`, 2018.

Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *arXiv preprint arXiv:2004.05439*, 2020.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2901–2910, 2017.

Charles Kemp and Joshua B Tenenbaum. The discovery of structural form. *Proceedings of the National Academy of Sciences*, 105(31):10687–10692, 2008.

Brenden Lake and Marco Baroni. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In *International Conference on Machine Learning*, pp. 2873–2882. PMLR, 2018.

Brenden M Lake. Compositional generalization through meta sequence-to-sequence learning. In *Advances in Neural Information Processing Systems*, pp. 9791–9801, 2019.

Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.

Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and brain sciences*, 40, 2017.

Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

Shirley Mark, Rani Moran, Thomas Parr, Steve W Kennerley, and Timothy EJ Behrens. Transferring structural knowledge across cognitive maps in humans and models. *Nature Communications*, 11 (1):1–12, 2020.

R Thomas McCoy, Erin Grant, Paul Smolensky, Thomas L Griffiths, and Tal Linzen. Universal linguistic inductive biases via meta-learning. *arXiv preprint arXiv:2006.16324*, 2020.

Anna Rogers, Olga Kovaleva, and Anna Rumshisky. A primer in bertology: What we know about how bert works. *arXiv preprint arXiv:2002.12327*, 2020.

Eric Schulz, Joshua B Tenenbaum, David Duvenaud, Maarten Speekenbrink, and Samuel J Gershman. Compositional inductive biases in function learning. *Cognitive psychology*, 99:44–79, 2017.

Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285, 2011.

Sebastian Thrun and Lorien Pratt. Learning to learn: Introduction and overview. In *Learning to learn*, pp. 3–17. Springer, 1998.

Jane X Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Remi Munos, Charles Blundell, Dharshan Kumaran, and Matt Botvinick. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*, 2016.

Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020.

ZD Zhang. Conjectures on the exact solution of three-dimensional (3d) simple orthorhombic ising lattices. *Philosophical Magazine*, 87(34):5309–5419, 2007.

A Appendix

A.1 Hyperparameter Details for Reinforcement Learning

We did a hyperparameter search for the following: value function coefficient, entropy coefficient, and learning rate. In particular, we evaluated each set of hyperparameters on a separate validation set, selected the model with the highest performing set, and re-trained the model to be evaluated on a previously-unseen test set of boards. Note that the final test set is not seen by the model until the last, final evaluation step. The different learning rates evaluated were: Searches were ran independently for both task distributions (compositional and null). The final selected hyperparameters for both task distribution were: value function coefficient=0.0006747109316677081,entropy coefficient=0.0006747109316677081, learning rate=0.0023483181861598565.

A.2 Code Availability

In our supplementary material as part of the submission, we've included code to generate compositional boards, train the network for the null task distribution, generate samples from the null task distribution via gibbs sampling, and train the meta-reinforcement learning agents used on the tile revealing task described in this work.