Biases towards compositionally simpler hypotheses are robust and unaffected by learning

Valerio Rubino (valerio.rubino@gmail.com)
CIMeC - Center for Mind/Brain Sciences, 101 Via delle Regole
Mattarello, TN 38068 38123, Italy

Peter Dayan (dayan@tue.mpg.de)
MPI for Biological Cybernetics
University of Tübingen Max-Planck-Ring 8, 72076 Tübingen, Germany

Charley M. Wu (charley.wu@uni-tuebingen.de)
University of Tübingen, Maria-von-Linden-Str. 6
Tübingen, Baden-Württemberg 72076 Germany

Abstract

Compositionality is an important and yet poorly understood feature of human cognition. In this study, participants navigated mazes with hidden, compositional structure, which was generated using operations over spatial primitives. Although they were not informed about the underlying structure, participants improved their accuracy and decreased primitive-inconsistent actions over the course of the task. Participants also selectively tested hypotheses corresponding to compositionally simpler expectations (simplicity bias), with a large proportion of errors due to expecting greater compositional structure than present in the true path. However, this simplicity bias did not change over the course of the experiment, but remained robust throughout. These results suggest that the human bias towards compositionality is unaffected by experience, at least in the timeframe of our experiment.

Keywords: compositionality, structure learning, simplicity, complexity, learning

Introduction

Compositionality has long been proposed as one of the most striking and singular features of human cognition (Chomsky, 2009; Dehaene, Al Roumi, Lakretz, Planton, & Sablé-Meyer, 2022; Kurth-Nelson et al., 2023), allowing us to understand novel and complex problems as abstract combinations of simpler components, bestowing remarkable flexibility and generalization. Although the mechanisms involved in compositional reasoning are still poorly understood, recent research in domains such as language (Hahn, Futrell, Levy, & Gibson, 2022), tool use (Thibault et al., 2021), concept learning (Goodman, Tenenbaum, Feldman, & Griffiths, 2008), visuospatial processing (Amalric et al., 2017; Schwartenbeck et al., 2021) and spatial navigation (Sharma, Curtis, Kryven, Tenenbaum, & Fiete, 2021) have re-ignited interest in this topic.

In particular, Kumar et al. (2022) showed that human participants are much better at uncovering hidden patterns that were generated using abstract, compositional operations

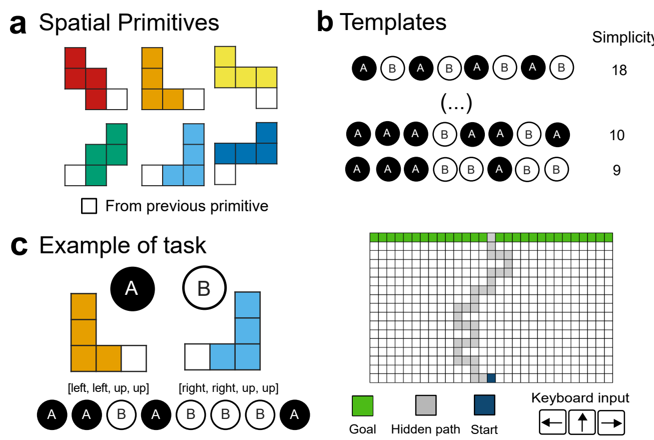

Figure 1: Task. **a**) The path of each maze was generated by drawing two primitives out of six possible ones. **b**) The two primitives of each maze were combined according to abstract templates, varying in their compositional simplicity. **c**) Participants navigated the task using the arrow keys and received feedback when committing mistakes.

(e.g., copy, symmetry, tree) compared to patterns matched in statistical complexity, but lacking compositionality. In contrast, neural networks displayed the opposite behavior, performing better on patterns without compositional structure. These results suggest people have priors favoring compositionality.

In the present study, we seek to understand the source of the human bias towards compositionality, by testing whether it is sensitive to experience. Drawing on previous research demonstrating human perception (auditory and visual) is modulated by the degree of compositional structure (Amalric et al., 2017; Planton et al., 2021; Sablé-Meyer et al., 2021), we use a spatial navigation task, where a single hidden path is generated using compositional operations over spatial primitives (Fig. 1). Rather than only remembering sequences (Amalric et al., 2017; Planton et al., 2021) or detecting differences (Sablé-Meyer et al., 2021), participants must iteratively test hypotheses about how to complete the maze. We use systematic patterns in hypothesis generation to measure their degree of bias favoring compositional structure, and examine how this changes as a function of experience.

Our results reveal that even though participants were not informed about the compositional structure of the mazes or the existence of spatial primitives, they learned to make more accurate and less primitive-inconsistent choices over the course of the experiment. However, experience did not influence their bias towards preferring hypotheses that corresponded to greater compositional structure (i.e., simplicity), which remained robust throughout the task.

Methods

Participants and Design. We recruited 59 participants from Prolific (30 female; M_{age}=37.32; SD=13.23). Participants were paid £2.80 for taking part in the experiment and a performance contingent bonus of up to £2.80. Participants spent 15.83 $\pm$ 0.72 minutes on the task and earned £4.32 $\pm$ 0.05 in total.

Materials and Procedure. After providing informed consent, participants were shown instructions, and completed an interactive tutorial and comprehension check before starting the experiment. The task consisted of navigating 20 mazes constructed on a 17 row by 33 column grid world, each with a single hidden solution that was generated with a compositional structure (see below). Each round began from the center tile on the bottom row, and the goal was to reach the top row colored in green (Fig. 1c). On each trial, participants chose an action using the left, up, or right arrow key to take a step. After each valid choice (moving within the grid and avoiding "backtracking"), participants received feedback, either progressing to the next tile or losing a life (from a total of 30 on each round) and remaining on the last correct tile. Participants were incentivised to complete each maze with as many lives left as possible, as this determined their bonus.

Each maze had a different correct path, defined by a compositional structure generated by sampling two out of a set of six possible primitives (Fig. 1a), each consisting of four actions (e.g., *up-left-left-up*). For each path, the two primitives were combined through an abstract template: for instance, the template *ABABABAB* specifies an alternation between primitive A and B, which repeats four times. Importantly, participants were never informed about the compositional structure of the task, nor the existence of primitives used in the generation of each path. Templates varied in their degree of compositional structure, which we defined using **simplicity**, which is a measure inspired by Alexander and Carey (1968), counting how often substrings (of any length) are "repeated" or "mirrored" in a sequence. Thus, more repetitions or mirroring of substrings correspond to higher simplicity and greater compositional structure. Previous work has found simplicity to be the best measure of performance in this type of task (Rubino, Hamidi, Dayan, & Wu, in press), compared against several alternative measures of compositional structure.

Results

Accuracy. Participants made increasingly more accurate choices over successive rounds (Odds Ratio (OR): 1.01 [1.01,1.01], $p < .001$; Fig. 2a). On average, participants selected correct actions with $P(\text{correct}) = .71$, beating baseline models using the true marginal ($P(\text{correct}) = .39$, $t(58) = 39.2$, $p < .001$) and true conditional probabilities of actions ($P(\text{correct}) = .44$, $t(58) = 32.9$, $p < .001$). This suggests participants learned to solve the mazes more efficiently with exposure, but do so by relying on more than just the (Markovian) statistics of the task.

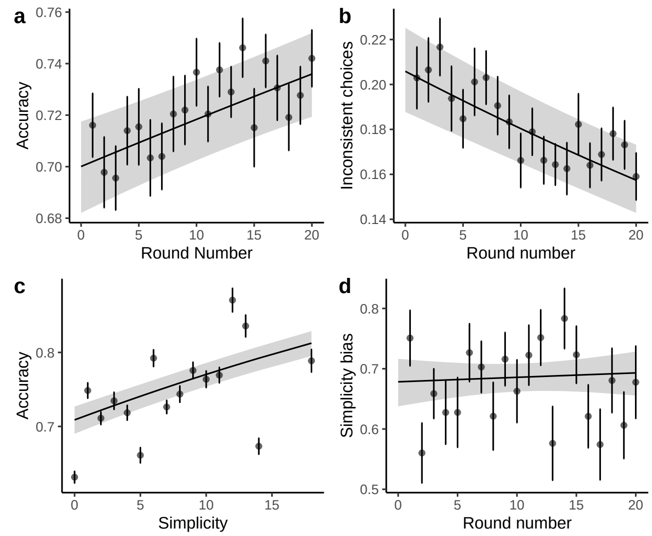

Figure 2: Results. **a**) Participants' accuracy increased over rounds. **b**) Even though participants were not informed about the existence of primitives or the underlying compositional structure, their proportion of primitive-inconsistent actions decreased over rounds. **c**) Participants were better at choosing actions corresponding to greater compositional structure (higher simplicity). **d**) Participants robustly tested hypotheses that corresponded to greater simplicity, which remained unchanged during the course of the task.

To measure the extent to which participants leveraged compositional structure, we computed the number of *primitive-inconsistent actions* (i.e., actions inconsistent with either of the two primitives used to generate the maze). Participants made inconsistent actions 19% of the time, which was significantly less than chance ($t(58) = -37.6$, $p < .001$). The proportion of inconsistent actions decreased with round number (OR: 0.98 [0.98,0.99], $p < .001$; Fig. 2b), suggesting participants refined their hypotheses to better match the primitives used to generate the mazes, even without explicit knowledge.

Simplicity. While a decrease in primitive-inconsistent actions suggest participant learned to detect and selectively deploy the primitives used in each maze, we also find evidence that participants combine primitives according to a *simplicity bias*, with actions favoring hypothesized subtemplates with greater compositional structure.

We first filtered for "distinguishing trials" where participants made primitive-consistent actions and when their possible choices corresponded to hypothesized subtemplate with differing simplicity. For instance, after observing the subtemplate ABA, the participant could choose either A or B to continue the sequence, yielding subtemplates with simplicity values of 2 and 3, respectively, making B correspond to a hypothesis with higher simplicity and thus greater compositional structure. We quantified simplicity bias as the proportion of trials in which participants chose an action corresponding to a simpler subtemplate vs. a more complex one.

Participants favored simpler subtemplates more consistently than chance (68%: $t(58) = 15.2$, $p < .001$, $d = 2.0$, $BF > 100$). Consequently, the simplicity of the true subtemplate influenced the accuracy of each choice, with participants making more accurate choices when they lead to higher simplicity (OR: 1.03 [1.03,1.04], $p < .001$; Fig. 2c). Indeed, 70% of times when participants chose a simpler subtemplate in distinguishing trials, they committed an error, due to expecting greater compositional structure than there actually was.

Strikingly, simplicity bias was not influenced by round number (OR: 1.00 [0.99,1.02], $p = .584$; Fig. 2d). And while we reported independent effects of round number and simplicity on accuracy, we find no significant interaction between them (OR: 1.00 [1.00,1.00], $p = .342$).

Conclusions

In this study, participants navigated mazes with hidden, compositional structure, which they were not explicitly informed about. The goal of the study was to examine whether a bias favoring compositionally simpler hypotheses would be influenced by experience. While accuracy and the selection of primitive-consistent actions improved over rounds, we find a robust and unchanging bias towards testing compositionally simpler hypotheses. Indeed, a large portion of errors corresponded to hypothesizing greater compositional structure than was present. These results shed light on the degree that a bias for compositionality adapts to experience, with further research needed over longer time-scales.

Acknowledgments

CMW is supported by 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 Germany's Excellence Strategy–EXC2064/1–390727645. PD is a member of the same cluster of excellence, and is funded by the Max Planck Society and the Humboldt Foundation. VR was funded by the Humboldt Foundation. We thank Philipp Schwartenbeck for insightful discussions and feedback on an early version of the experiment.

References

Alexander, C., & Carey, S. (1968). Subsymmetries. *Perception & Psychophysics*, *4*(2), 73–77.

Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., & Dehaene, S. (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS computational biology*, *13*(1), e1005273.

Chomsky, N. (2009). *Cartesian linguistics: A chapter in the history of rationalist thought*. Cambridge University Press.

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences*.

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.

Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, *119*(43), e2122602119.

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

Kurth-Nelson, Z., Behrens, T., Wayne, G., Miller, K., Luettgau, L., Dolan, R., ... Schwartenbeck, P. (2023). Replay and compositional computation. *Neuron*.

Planton, S., van Kerkoerle, T., Abbih, L., Maheu, M., Meyniel, F., Sigman, M., ... Dehaene, S. (2021). A theory of memory for binary sequences: Evidence for a mental compression algorithm in humans. *PLoS computational biology*, *17*(1), e1008598.

Rubino, V., Hamidi, M., Dayan, P., & Wu, C. M. (in press). Compositionality under time pressure. In M. Goldwater, F. Anggoro, B. Hayes, & D. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.* Sydney, Australia: Cognitive Science Society.

Sablé-Meyer, M., Fagot, J., Caparos, S., van Kerkoerle, T., Amalric, M., & Dehaene, S. (2021). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. *Proceedings of the National Academy of Sciences*, *118*(16), e2023123118.

Schwartenbeck, P., Baram, A., Liu, Y., Mark, S., Muller, T., Dolan, R., ... Behrens, T. (2021). Generative replay for compositional visual understanding in the prefrontal-hippocampal circuit. *bioRxiv*.

Sharma, S., Curtis, A., Kryven, M., Tenenbaum, J., & Fiete, I. (2021). Map induction: Compositional spatial submap learning for efficient exploration in novel environments. *arXiv preprint arXiv:2110.12301*.

Thibault, S., Py, R., Gervasi, A. M., Salemme, R., Koun, E., Lövden, M., ... Brozzoli, C. (2021). Tool use and language share syntactic processes and neural patterns in the basal ganglia. *Science*, *374*(6569), eabe0874.