Learning Dynamics of Semantic Knowledge in Humans and Neural Networks

Jirko Rubruck (jirko.rubruck@psy.ox.ac.uk)
Department of Experimental Psychology, University of Oxford, Oxford, UK

Andrew Saxe (a.saxe@ucl.ac.uk)
Gatsby Unit and Sainsbury Wellcome Centre, University College London, London UK;
CIFAR Azrieli Global Scholars Program, CIFAR, Toronto, Canada

Christopher Summerfield (christopher.summerfield@psy.ox.ac.uk)
Department of Experimental Psychology, University of Oxford, Oxford, UK

Abstract

Semantic cognition allows for flexible acquisition, storage, and deployment of conceptual representations. Previously, different computational models have attempted to explain a wealth of empirical phenomena in semantic learning (Rogers & McClelland, 2004; Saxe et al., 2019). Our work focuses on phenomena documented in the form of analytical solutions to learning dynamics of deep linear networks which display progressive differentiation and stage-like transitions during learning. These phenomena have been frequently observed during human development, but have not been tested in the formal setting of a controlled experiment. Here, we asked if they are general features of human learning that persist when training adults over short time-spans. We trained participants on a semantic learning experiment that invoked hierarchical constraints on learned properties. Given our theoretical predictions, we compared human data to several classes of differently initialised neural networks. Findings indicate that human learning respects the hierarchical constraints. Furthermore, we find that human learning is most closely mirrored by neural networks which learn from small rather than large random weights. Such networks in particular are known to display patterns of progressive differentiation and stage-like transitions. Our results validate that simple neural networks can usefully describe phenomena observed in human semantic cognition.

Keywords: Semantic learning; Learning dynamics; Deep linear networks;

Introduction

In humans the acquisition, storage, and retrieval of semantic information appears to be subject to constraints and empirical regularities. Semantic information about the real world appears to obey hierarchical constraints (Rogers & McClelland, 2004). Furthermore, the acquisition of semantic information during human development is argued to proceed in structured, progressive fashion whereby more general high-level semantic distinctions are learned before fine-grained, object specific information (Keil, 2013). The acquisition of hierarchical semantic distinctions have also been found to obey non-linear dynamics in developing children whereby abrupt improvements in conceptual knowledge signal the acquisition of new levels in taxonomical semantic structures. We examine if the phenomena of progressive differentiation and stage-like transitions can also be observed in a formal lab experiment with adult human participants.

Models of Semantic knowledge

In the current work, we focus on results obtained in the study of deep linear networks. This simplified model class omits non-linear activation functions. In contrast to their non-linear counterparts such simplified models allow for exact, analytical solutions to learning dynamics captured in the evolution of model weights. Despite the loss of expressivity of functions in the hypothesis space of linear models, they allow for the modelling of many key phenomena prevalent in semantic cognition in a mathematically exact fashion. Furthermore, learning in such models will often resemble the learning dynamics of non-linear models closely (Saxe et al., 2019).

Methods

Experimental Procedure We trained a cohort of *N = 49* adult participants on a semantic learning task requiring them to learn the mapping from a set of visual stimuli to their corresponding semantic properties. The learned properties contained an implicit hierarchical structure. On each trial, participants were presented with the visual stimulus and buttons representing semantic properties. Participants selected a subset of three buttons and received fully informative feedback.

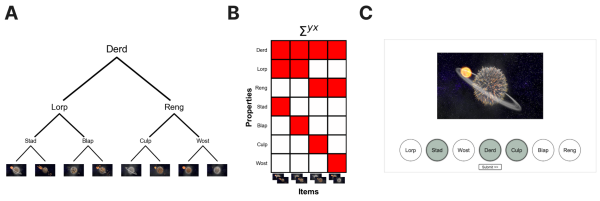

Figure 1: (A) Hierarchical dataset used in the semantic learning task. Properties of each item correspond to a walk through the graph starting at a leaf and ending at the root node. (B) Dataset input-output correlation matrix. (C) Example screen from a trial presented to participants during the learning task.

Model simulations To assess the correspondence of our participant responses with results obtained in neural networks, we trained different classes of simple neural networks on an analogous semantic learning task. The dataset used for the model simulations consisted of $P = 4$ input-output pairs $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{P}$. Inputs in the task were whitened such that $\mathbf{X} = \mathbf{I} \in \mathbb{R}^{4\times4}$. Labels $\mathbf{Y} \in \mathbb{R}^{7\times4}$ contained a hierarchical structure. The input-output correlation matrix of the dataset $\mathbf{\Sigma}^{yx} = \frac{1}{P}\sum_{i=1}^{P} \mathbf{y}_i\mathbf{x}_i^T$ hence takes the form shown in Fig. 1B. We trained two types of networks on the semantic learning task predicting properties from input objects as $\hat{\mathbf{y}}_i = \mathbf{W}^2\mathbf{W}^1\mathbf{x}_i$ and $\hat{\mathbf{y}}_i = \mathbf{W}^2 ReLU(\mathbf{W}^1\mathbf{x}_i)$ with $\mathbf{W}^1 \in \mathbb{R}^{16\times4}$ and $\mathbf{W}^2 \in \mathbb{R}^{7\times16}$. We initialised model weights as $\mathbf{W}^1_{ij} \sim \mathcal{N}(0, a_0^2/N_1)$, $\mathbf{W}^2_{ij} \sim \mathcal{N}(0, a_0^2/N_3)$. We parameterised these distributions with $a_0^2 \in \{1e0, 1e-1, 1e-2, 1e-3, 1e-4, 1e-5\}$ to examine correspondence of initialisation related changes in learning dynamics to human learning.

Results

Human Learning

We can observe that participant learning respects the hierarchical structure which underlies the presented semantic properties. We find that participants learn the top level in the hierarchy at a faster pace and can perform with nearly perfect ac-

curacy in later blocks Fig. 2A. The difference in performance between the lowest and the mid-level of our hierarchy is especially remarkable as the lowest properties are most specific to individual classes of planets and are therefore a more unique identifier. The results are in line with our predicted progressive differentiation during semantic learning.

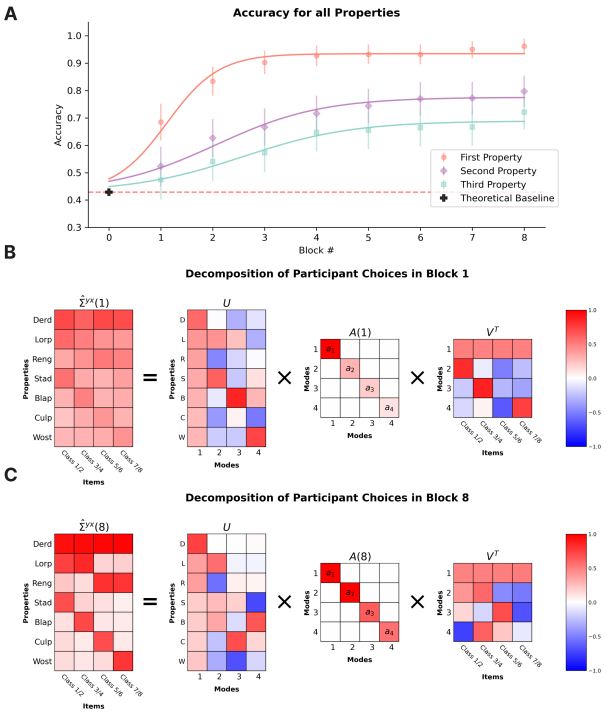

Figure 2: (A) Accuracy of participants across all 3 property levels. Error bars indicate SEM. The red dotted line indicates chance. (B) SVD of human choices during initial stage of learning. The large singular value a_1 on the Diagonal of $A(0)$ reflects performance on the highest level of the hierarchy. (C) SVD of our participant choices at the end of learning. Hierarchy of singular values on the diagonal recapitulate dataset eigenstructure.

A two-way repeated measures ANOVA confirmed this qualitative observation and revealed a significant main effects of block $F(7,336) = 41.635$, $p < 10^{-4}$, $\eta^2 = 0.151$ and level $F(2,96) = 98.87$, $p < 10^{-4}$, $\eta^2 = 0.26$ on participants' accuracy on the semantic learning task. However, the interaction between block and level was not significant $F(14,672) = 0.251$, $p = 6.74 \times 10^{-2}$, $\eta^2 = 0.005$.

Human input-output correlation matrices

To understand the patterns of progressive hierarchical differentiation and stage-like transitions, we assess participants' time dependent input-output correlation matrices $\hat{\mathbf{\Sigma}}^{yx}(t)$ similar to Saxe et al. (2019). We can break down this input-output correlation matrix using singular value decomposition (SVD) which generalises the spectral decomposition of symmetric matrices to the non-symmetric case. The decomposition of the input-output matrix appears similar to the deep linear networks by Saxe et al. (2019) towards the last block of human learning Fig 2C. In addition, we find that the large singular value a_1 in the beginning of learning recapitulates stronger knowledge of high-level semantic properties in early stages of learning and the yet incomplete acquisition of dataset eigenstucture by our human participants Fig 2B.

Comparing Human and Network Input-Output correlation matrices

To further understand the pattern of progressive differentiation and stage-like transitions we compare our obtained human input-output correlation matrices with neural networks of different initialisation schemes and architectures. Patterns of progressive differentiation and stage-like transitions are seen as a results of training networks from small initial weights where $a_0^2 \ll 1$. We averaged over input-output matrices in 8 equal partitions of training epochs to gain input-output correlation matrices comparable to those obtained from human participants across 8 blocks of learning. To assess the similarity of these network input-output correlation matrices to our human data, we calculated the Euclidean distance between vectorised human and network matrices for our Relu and Linear networks as shown in Fig. 3.

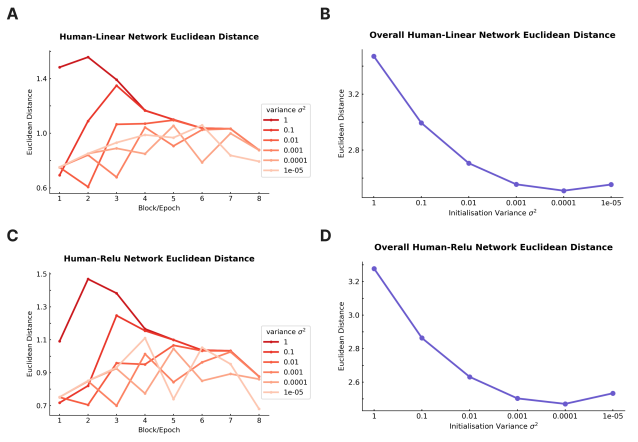

Figure 3: Euclidean distance between vectorised input-output-correlation matrices of humans and neural network input-output correlation matrices. (A) and (C) display distance as a function of block and initialisation variance. (B) and (D) display the overall distance between humans and neural network matrices as a function of initialisation variance.

Euclidean distances between networks and humans are smallest when training networks from small initial weights where $a_0^2 \ll 1$. The result supports our prediction that human learning in response to hierarchically structured semantic data displays progressive differentiation and stage like transitions.

Acknowledgments

This work was supported by the European Research Council (award REP-725937 to C.S.) and through a ESRC GUDTP doctoral stipend (awarded to J.R.).

References

Keil, F. C. (2013). Semantic and conceptual development. In *Semantic and conceptual development*. Harvard University Press.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2019). A mathematical theory of semantic development in deep neural networks. *Proceedings of the National Academy of Sciences*, *116*(23), 11537–11546.