The Effects of Learning on the Representational Geometry of Skilled Chess Players

Andrea Ivan Costantino (andreaivan.costantino@kuleuven.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Esna Mualla Gunay (esnamualla.gunay@kuleuven.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Emily Van Hove (emily.vanhove@student.kuleuven.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Laura Van Hove (laura.vanhove@kuleuven.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Felipe Fontana Vieira (felipe.fontanavieira@ugent.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Merim Bilalic (merim.bilalic@northumbria.ac.uk)
Department of Psychology, Northumbria University,
Newcastle Upon Tyne, UK

Hans Op de Beeck (hans.opdebeeck@kuleuven.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Abstract

Chess, with its rich history as a metaphor for human intelligence, offers an excellent framework to examine expertise effects. Previous studies suggested that expert players analyse chess boards differently from novices, emphasizing piece relationships over visual traits. However, these studies did not explore representational structure and information processing changes in expertise, and in what brain areas these changes may occur. Our work bridges this gap by employing computational, behavioural, and neuroimaging methodologies to uncover representational changes in expert biological and artificial systems. By comparing chess expert and non-expert systems in humans (fMRI) and *in silico* (DNNs), we aim to identify chess expertise's representational changes. Our results reveal similar information processing between humans and DNNs, showing a representational and behavioural alignment between expert systems. Additionally, experts systems show a representational reorganization, resulting in more linearly separable representations of relevant high-level dimensions in late processing stages.

Keywords: learning; expertise; neural networks; neuroimaging

Introduction

Expertise, defined as the ability to consistently achieve superior performance in a specific domain, manifests through profound behavioural and representational changes. Previous studies on expertise demonstrate that chess experts process game-related information more effectively than novices. They possess faster and more accurate pattern recognition abilities, display fewer but more strategically targeted visual fixations, and exhibit increased activations in brain regions involved in visual recognition and strategic planning (Bilalić, Langner, Ulrich, & Grodd, 2011; Reingold, Charness, Pomplun, & Stampe, 2001). Previous studies on learning have already to some degree investigated changes in information processing, but despite these insights there is still only limited knowledge of how profoundly expertise can change information processing and internal representations at the neural and the computational level of both human and artificial intelligent systems. In this context, chess offers unique opportunities considering the wealth of data about changes at the cognitive level and the complexity of the underlying information processing. Our study seeks to fill this gap by investigating how chess expertise alters information processing in humans and DNNs. We hypothesize (i) that artificial and human experts will exhibit distinct neural patterns for chess-relevant features in high-level regions, reflecting their advanced game understanding; and (ii) a degree of behavioural and representational alignment between expert systems. By characterizing the representational changes observed in our participants, our study strives to deepen our understanding of the organizing principles of learning and expertise.

Results

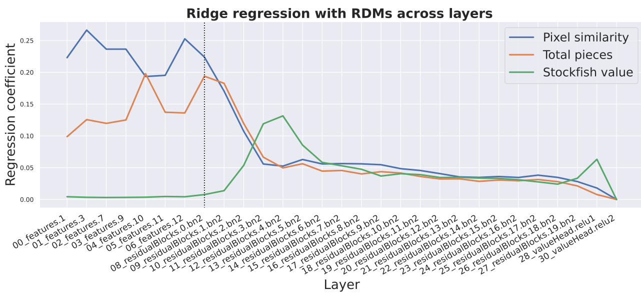

Figure 1: Regression coefficients across DNN layers. The dotted line marks the end of the visual stream and the start of the chess processing stream.

Chess information processing is hierarchical We developed a two-stages deep neural network (DNN) comprising a *visual* processing stream and a *chess* processing stream. The visual stream employs a modified AlexNet (Krizhevsky, 2014) architecture trained to convert chessboard images into structured board representations, while the chess stream uses a publicly available AlphaZero (Silver et al., 2018) implementation trained with supervised learning algorithms. These two streams were trained independently on distinct tasks before being integrated for final testing. This structure allows for a nuanced comparison with neuro-imaging data.

We tested the network on a dataset of 5,000 chessboard images. This dataset was designed to vary along both low-level (e.g., pixel similarity, total number of pieces) and high-level (e.g., predicted move value by Stockfish) dimensions. To analyze these activations, we constructed Representational Dissimilarity Matrices (RDMs) for each layer and tested dimension (pixel similarity, total number of pieces, and Stockfish value). These RDMs were then used in ridge regression analyses to determine how each layer's information processing related to the different dimensions of board analysis. The coefficients from these regressions are displayed in Figure 1.

The findings underscore a hierarchical organization of information within the network, where initial layers predominantly handle low-level visual details, and higher layers increasingly engage with complex chess strategies, in a progression that is consistent with our understanding of human cognitive processes.

Task-relevant information is encoded by experts We conducted an fMRI study involving 20 chess experts and 13 novices. Participants performed a 1-back task on a set of 40 chessboards that varied along specific dimensions: visual similarity, presence of checkmate, and the type of checkmate based on the piece combination. Participants were asked to choose the more strategically advantageous position between two consecutive chess boards.

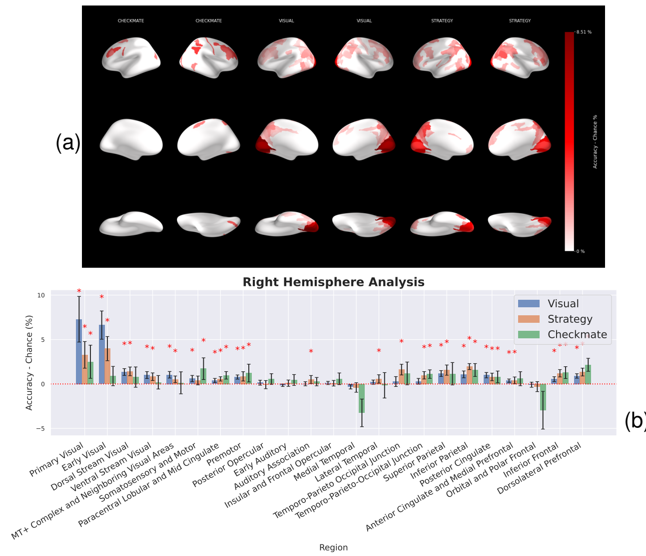

Figure 2: SVM Decoding Results Across HCP-MMP1 ROIs: (a) Accuracy above chance (%) on *fsaverage* cortical surface, showing only significant ROIs at $p < 0.01$ (FDR corrected). (b) Accuracy in right hemisphere groups of ROIs, arranged posterior to anterior.

After minimal preprocessing, fMRI data was modelled using trials as regressors and masked using the HCP-MM1 parcellation (Glasser et al., 2016) projected into participants' MNI volumes. Low-level visual features and high-level strategic concepts ('Visual', 'Strategy', 'Checkmate') were then decoded from the resulting beta images using multivariate decoding techniques (see Figure 2).

The analysis indicates that high-level representations in chess experts' brains are linearly separable within higher regions. Similar to the hierarchical information processing observed in DNNs, the neural architecture in humans also exhibits a hierarchical pattern: posterior brain regions predominantly encode low-level visual information, anterior regions process high-level strategic information, and the occipito-temporal-parietal clusters integrate aspects of checkmate type, reflecting a combination of perceptual and strategic processing. This suggests a parallel in the way that both human brains and artificial networks manage complex visual and cognitive tasks, and identifies the brain location of chess expertise effects.

Behavioural similarity In this analysis, we explore the error patterns of human participants during an online familiarization task, comparing these patterns with those of two chess models: the Lc0 (The LCZero Authors, n.d.) model, which is trained using reinforcement learning algorithms and achieved close-to-ceiling performance on the previously described 5000

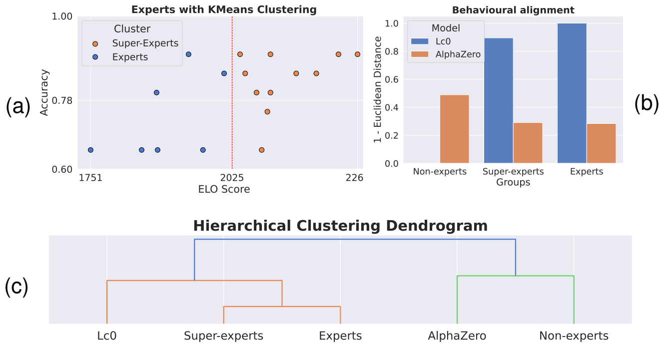

Figure 3: Behavioural alignment results: (a) KMeans clustering of experts by ELO and performance; (b) Error pattern comparisons between groups and models; (c) Hierarchical clustering dendrogram.

boards dataset, and a supervised version of AlphaZero, which was used in previous analyses and achieved an overall lower performance on the same dataset ($\mu_{\text{accuracy}} = 0.58$). Both the human participants and the models were evaluated on the same 40-board dataset employed in our prior fMRI study.

To further differentiate levels of expertise among human participants, we applied K-means clustering to their accuracy scores and Elo ratings, distinguishing between 'super-experts' and 'experts'. We computed the similarity between the error patterns of each human group (super-experts, experts, non-experts) and those generated by each model, using *1 - Euclidean distance* as similarity measure. The distance between humans' and DDNs' pattern of error was used here as a benchmark for behavioural alignment, following other recent studies highlighting the effectiveness of this measure (Maniquet, Op de Beeck, & Costantino, 2024).

The results (see Figure 3) demonstrated a stronger behavioural alignment between the expert groups and the Lc0 model. In contrast, the non-expert group showed a higher alignment with AlphaZero. These findings indicate that extended periods of unsupervised training might be essential for achieving higher behavioural congruence between expert models and human behaviour. Additionally, the results corroborate the effectiveness of our dataset in distinguishing between expert and non-expert chess players.

Conclusions

Overall, our results deepen our understanding of learning in human and artificial intelligent systems, and draw new parallels between humans' and DNNs' internal processes.

Acknowledgments

We thank Dr. Artem Platonov (plartem@gmail.com) for his early feedback and contributions to the fMRI dataset. This work was funded by the following grants awarded to H.O.B.: FWO project on expertise G0D3322N, KU Leuven C1 project C14/21/047, Methusalem project METH/24/003.

References

Bilalić, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: fusiform face area in chess experts and novices. *Journal of neuroscience*, *31*(28).

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... others (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.

Krizhevsky, A. (2014). One weird trick for parallelizing convolutional neural networks. *arXiv preprint arXiv:1404.5997*.

Maniquet, T., Op de Beeck, H., & Costantino, A. I. (2024). Recurrent issues with deep neural network models of visual recognition. *bioRxiv*. doi: 10.1101/2024.04.02.587669

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological science*, *12*(1).

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... others (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, *362*(6419), 1140–1144.

The LCZero Authors. (n.d.). *LeelaChessZero.* Retrieved from `https://github.com/LeelaChessZero/lc0`