Investigating the Emergence of Complexity from the Dimensional Structure of Mental Representations

Trisha Mazumdar (trisha.mazumdar@vanderbilt.edu)
Department of Computer Science, Vanderbilt University

Vivian Cai (vivian.cai@vanderbilt.edu)
Department of Computer Science, Vanderbilt University

Stephen Chong Zhao (chong.zhao.1@vanderbilt.edu)
Data Science Institute, Vanderbilt University

Jason Lee (jason.j.lee@vanderbilt.edu)
Department of Computer Science, Vanderbilt University

Andrew Bender (abender@health.ucsd.edu)
Neurosciences Graduate Program, University of California San Diego

Adaline Leong (jia.yin.leong@vanderbilt.edu)
Department of Computer Science, Vanderbilt University

Jad El Harake (jad.el.harake@vanderbilt.edu)
Biomedical Engineering Department, Vanderbilt University

Mark Wallace (mark.wallace@vanderbilt.edu)
Department of Psychology, Vanderbilt University

David A. Tovar (david.tovar@vanderbilt.edu)
Department of Psychology, Vanderbilt University

Abstract

Objects can be described by various dimensions that, when combined, form a distinct entity. This study explores the multi-dimensional structure of mental representations of objects and the associated property of complexity. We investigated how different characteristics influence perceived complexity and evaluated the predictive power of entropy scores as indicators of this complexity. Our results show that entropy scores, calculated from mental embeddings and adjusted by perceptual weights, can predict perceptual complexity effectively. Notably, once these weights are tuned to the relative complexity of each dimension, entropy scores based on human complexity ratings significantly enhance the correlation between entropy and participant choices in distinguishing between ambiguous and control images. Importantly, we established a complexity score using a perceptually tuned CLIP model, CLIP-HBA, that makes this metric generalizable to novel stimuli due to its ability to detect perceptually relevant dimensions in objects.

Keywords: mental embeddings, object recognition, complexity, entropy, ambiguity

Introduction

The brain must synthesize many different aspects of an object in order to discern its identity or meaning. The mental construction process involves determining which categories an object might fall into based on the aspects we perceive. Then, our final identification of the object relies on the intersection of and interaction between those categories (Zheng et al., 2019; Kriegeskorte and Diedrichsen, 2019). In a previous study, Hebart et al. (2020) identified up to 49 distinct "dimensions" (e.g. animal-related, metal-related, thin/flat, etc.) that can be used to predict human similarity judgements. They hypothesized that these dimensions encompass the overall structure of our mental representations of objects. Given that not all objects can be characterized by the same number of dimensions, and the interplay between dimensions will differ based on what they are, we raise a question regarding an emergent property of these intricacies: what makes an object complex? Emergent properties are characteristics of a system that arise from the interaction and interrelation of its parts, rather than from the individual components themselves. One thought of complexity is that it arises from the constituent parts that must be integrated together within a system, in theories such as Information Integration Theory (IIT) (Oizumi et al., 2014). Building on this foundation, we explored the application of entropy scores derived from the THINGS dataset embeddings (Hebart et al., 2020) as predictors of complexity. Entropy is related to the uncertainty or unpredictability of a system's state (Jia and Wang, 2024). For this study, we calculated entropy using the THINGS image embeddings from the work of Hebart et al. (2020) and a version of CLIP trained on the THINGS dervied mental embeddings, called CLIP-HBA (Zhao, under review). Images that exhibited a wider dispersal of values across the 49 dimensions were calculated to have higher entropy scores. We hypothesize that there exists a positive relationship between the number of dimensions which characterize an object and complexity, since more dimensions leads to more potential interactions and relationships within the conceptual understanding of an object. Given the possibility that some dimensions may be more complex than others, we asked subjects to rank images and calculated dimensional weights based on the results. Then, we used these weights to calculate and recalibrate the entropy of ambiguous images and style-control images passed through CLIP-HBA.

Complexity Ranking of THINGS Images

We first used the 1854 original THINGS embeddings (Hebart et al., 2020) to calculate entropy scores for each image. Each embedding was normalized and turned into a probability distribution. Then, we measured the entropy of each distribution to obtain a numerical entropy score which described the "spread" of dimensions contained within each image. Figure 1B demonstrates the contrast between a high-entropy and a low-entropy image. The rose plot of the image of squid displays fewer dimensions than the rose plot of the image of a watch. Subsequently, we categorized the images into five groups from low to high entropy (Figure 1C), and randomly selected 10 images from each group, totaling 50 images. We then presented these images to 14 participants who ranked them from most to least complex (Figure 1D). These responses then allowed us to calibrate entropy based on participant perceptions of dimension complexity, identifying which dimensions contributed more significantly to the overall complexity. We then tested these calibrated dimension scores on an ambiguous object task

Complexity Comparisons of Ambiguous Images

We first obtained 36 ambiguous images and created 36 style-transferred control images through stable diffusion models (Kalra, 2024) and Adobe Firefly. Figure 1E displays an example of an ambiguous image vs. its style-transferred control. To determine whether the ambiguous images exhibited a higher level of complexity than the control images, we displayed each ambiguous image along with its control and asked participants to choose which one was more complex. From their responses, we calculated the difference between their responses as scores— positive scores indicating a higher preference for ambiguous images, and negative scores indicating a higher preference for control images. Then, the 36 ambiguous images and their controls were run through CLIP-HBA (Zhao, under review). The resultant embeddings were turned into entropy scores and re-weighted for dimensional complexity using the com-

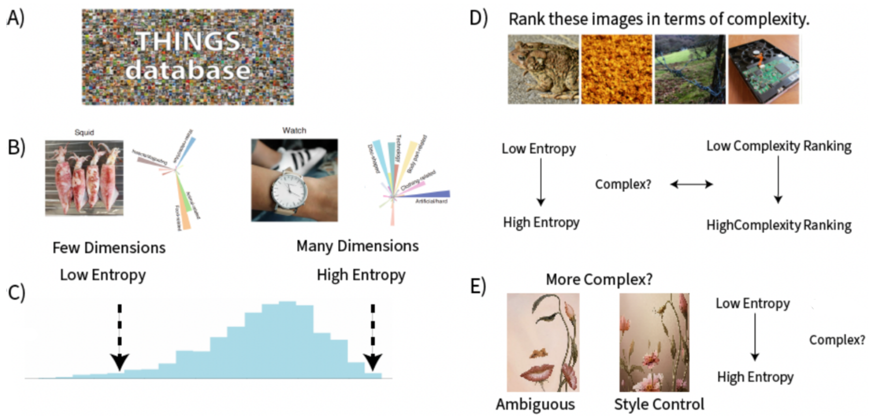

Figure 1: Experiment Schematic

plexity score from the THINGS image ranking experiment described above. We performed 3 different Spearman correlations with the participant scores: one with the CLIP-HBA entropy scores before weighting, one with CLIP-HBA entropy scores after weighting, and one with the output from another image quality assessment model, CLIP-IQA (Wang et al., 2022). For all of our correlations, we first calculated the pairwise differences between the average proportions of participant ambiguous choice and participant control choice. Then, we calculated the corresponding pairwise proportional differences of unweighted CLIP-HBA entropy, weighted CLIP-HBA entropy, and CLIP-IQA complexity to correlate the participant data with. We found the Spearman correlation coefficient between original participant scores before reweighting, and the complexity scores from CLIP-IQA yielded a coefficient of 0.524 ($p = 0.001$). The Spearman correlation coefficient between original participant scores and the entropy scores from CLIP-HBA was -0.488 ($p = 0.003$) due to a greater shift towards the ambiguous choice over the control choice. Finally, the correlation between the participant results and the differences between proportions of weighted CLIP-HBA entropy scores showed a significant improvement over both the comparison with unweighted CLIP-HBA entropy scores and the comparison with CLIP-IQA complexity values $(\rho = 0.8350, p = 4.4708^{-10})$ (see Figure 2). This suggests that the weights obtained from the analysis of the Hebart embeddings in the context of our ranking experiment might be generalizable to external image datasets, since they came from participant data in the context of a completely different set of images. Additionally, the significant improvement in correlation between the unweighted CLIP-HBA entropy scores and the weighted CLIP-HBA entropy scores suggests a combination of entropy scores from CLIP-HBA and individual dimensional complexity is an indicator of human perceptions of complexity. In future studies, neural component could be incorporated when viewing the images. Additionally, the control images used by the participants were generated in a style that did not consistently eliminate the dual-meaning quality. Instead, the diffusion model added finer details to the images, such as skin textures, flowers, and trees in the background. The 36 ambiguous images were also selected arbitrarily; further studies should explore more images in a wider range of categories.

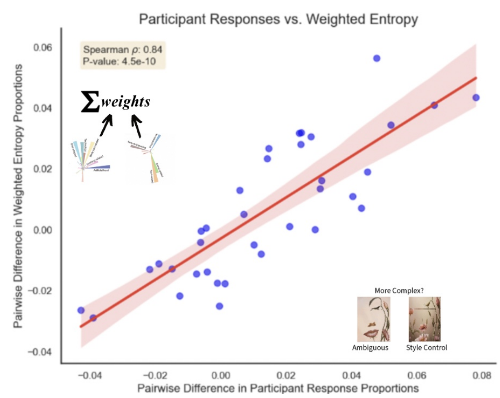

Figure 2: Participant Responses vs. Weighted Entropy

Conclusion

The results of our study and subsequent analysis provide new insights regarding the emergent properties of complexity in mental representations. We demonstrate that a combination of metrics, including entropy along with assessing the complexity that may be contained within dimensions, provide a better understanding of nuanced human perceptions of complexity than any standalone metric. In future work, we aim to expand this methodology to a broader range of images and refine our analyses based on additional human cognitive data, enhancing the generalizability and accuracy of complexity predictions.

Acknowledgements

This work was supported by Vanderbilt's IDD Reads Grant.

References

Hebart, M. N., Zheng, C. Y., Pereira, F., and Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, 4(11):1173–1185.

Jia, H. and Wang, L. (2024). Introducing entropy into organizational psychology: An entropy-based proactive control model. *Behavioral Sciences*, 14(1):54.

Kriegeskorte, N. and Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience*, 42:407–432.

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Computational Biology*, 10(5):e1003588.

Wang, J., Chan, K. C. K., and Loy, C. C. (2022). Exploring clip for assessing the look and feel of images.

Zheng, C. Y., Pereira, F., Baker, C. I., and Hebart, M. N. (2019). Revealing interpretable object representations from human behavior.