Five animacy dimensions and the CLIP model explain complementary components of visual representational dynamics and similarity judgments

Kamila Maria Jozwik (jozwik.kamila@gmail.com)
Department of Psychology, University of Cambridge, Downing Site, Craik-Marshall Building
Cambridge, United Kingdom

Radoslaw Martin Cichy (rmcichy@googlemail.com)
Department of Education and Psychology, Freie Universität Berlin
Berlin, Germany

Nikolaus Kriegeskorte (nk2765@columbia.edu)
Zuckerman Mind Brain Behavior Institute, Department of Psychology, Department of Neuroscience, Columbia University
New York, USA
These authors jointly supervised this work: Radoslaw M. Cichy, Nikolaus Kriegeskorte

Abstract

Distinguishing animate from inanimate things is important for object recognition behaviour and animate and inanimate objects elicit distinct brain and behavioural responses. A recent study evaluated the importance of five object dimensions related to animacy ("being alive", "looking like an animal", "having agency", "having mobility", and "being unpredictable") in brain representations and similarity-judgement behaviour. The study introduced a stimulus set that decorrelated these dimensions based on human ratings. Here, we ask: 1) to what extent one of the best computational models of vision (Contrastive Language-Image Pre-Training (CLIP) RN50) can predict dynamic human brain (EEG) and similarity judgement responses to this stimulus set and 2) what unique variance is explained by each animacy dimension ratings and CLIP. We find that CLIP explains a unique portion of the variance of similarity judgements, and a similar total amount of the variance as human ratings for each of the animacy dimensions. EEG responses are also predicted by animacy dimension ratings and CLIP to a similar extent. However, CLIP explains a unique portion of this variance at short latency (140-196 ms after stimulus onset), whereas "looking like animal" dimension rating explains unique variance at longer latency (239-301 ms after stimulus onset). We conclude that both human-generated multi-dimensional animacy ratings and the CLIP model explain unique components of visual representational dynamics and similarity-judgement behaviour and provide insights about specific dimensions of animacy that need to be better captured in future computational models of brain function and behaviour.

Introduction

Animate and inanimate objects elicit distinct brain (Kriegeskorte et al., 2008; Cichy, Pantazis, & Oliva, 2014) and behavioural (Mur et al., 2013) responses. Five object dimensions related to animacy were reported in the literature: "being alive" (Connolly et al., 2012), "looking like an animal" (Bracci, Ritchie, Kalfas, & Op de Beeck, 2019), "having agency" (Thorat, Proklova, & Peelen, 2019), "having mobility" (Beauchamp, Lee, Haxby, & Martin, 2002), and "being unpredictable" (Lowder & Gordon, 2015)). A recent study created a stimulus set decorrelating these dimensions as much as possible based on each dimension human ratings, and evaluated the importance of each of these dimensions in brain and behaviour (Jozwik et al., 2022). Here we extend this work by modelling. First, we wanted to know to what extent one of the best (Conwell, Prince, Kay, Alvarez, & Konkle, 2024) computational models of vision (Contrastive Language-Image Pre-Training (CLIP) RN50 (Radford et al., 2021)) can predict dynamic brain (EEG) and behavioural (similarity judgements) responses to this carefully designed stimulus set (Figure 1). Secondly, we asked what unique variance is explained by each animacy dimension rating and CLIP to understand what information about visual brain and behavioural representations may be missing in computational models and behavioural animacy ratings.

Figure 1: The genetic-algorithm driven stimulus set consisted of 128 images decorrelated on five dimensions of animacy.

Methods and Results

The same 19 subjects participated in EEG (stimulus presentation duration: 500 ms, inter-trial interval: 1–1.1 s) and similarity judgements experiments (details of the stimulus generation and experimental design can be found in Jozwik et al. (2022)). We computed response patterns (across animacy dimension ratings, similarity judgements, EEG signals, and CLIP layers) for each image. We then computed response-pattern dissimilarities between images (using Euclidean distance as a metric) and placed these in a representational dissimilarity matrix (RDM). We correlated each animacy dimension rating and the best CLIP layer (for the EEG and similarity judgements data it is visual layer 4) with the data. We find that CLIP explains similar amount of variance in similarity judgements as compared to each animacy dimension rating (Figure 2) and explains unique portion of this variance (Figure 3). EEG responses are also predicted by animacy dimension ratings and CLIP to a similar extent (Figure 4), however CLIP explains unique portion of this variance at short latency (140-196 ms after stimulus onset), whereas "looking like animal" dimension rating explains unique variance at longer latency (239-301 ms after stimulus onset, Figure 5). We conclude that both human-generated animacy ratings and computational model explain unique components of visual representational dynamics and behaviour.

Discussion

Our findings are consistent with previous work showing that human-generated labels and DNNs explain unique variance in source-reconstructed MEG data (Jozwik, Kietzmann, Cichy, Kriegeskorte, & Mur, 2023). The high performance of CLIP here is consistent with it predicting brain representations well (Conwell et al., 2024) using the 7T fMRI Natural Scenes Dataset (NSD, Allen et al. (2022)). An interesting future avenue is to test whether findings based on the unique

and carefully designed dataset used in this study would generalize to larger publicly available datasets such as NSD or THINGS (Hebart et al., 2023). Finally, these results provide insights about specific dimensions of animacy that need to be better captured in future computational models of brain function and behaviour.

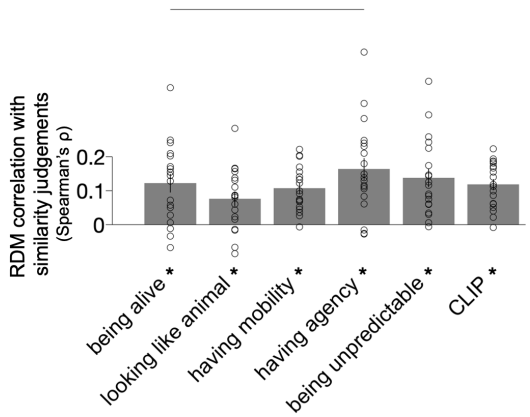

Figure 2: Animacy dimension RDM comparisons with similarity judgements RDMs (significant correlation - asterisk (one-sided Wilcoxon signed-rank test, $p < 0.05$ corrected), error bars - the standard error of the mean based on single-participant correlations, circles - single-participant correlations, horizontal lines - significant pairwise differences between model performance ($p < 0.05$, FDR corrected across all comparisons).

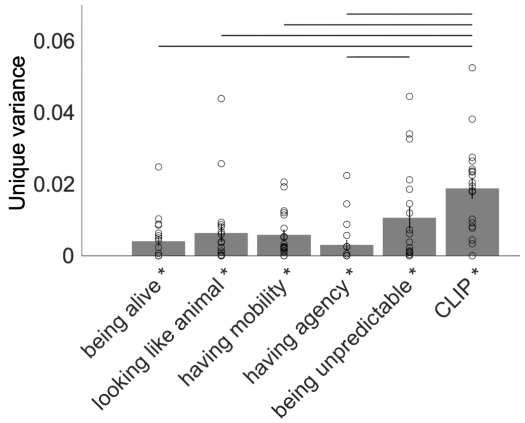

Figure 3: Unique variance of each animacy dimension in explaining similarity judgements using the same visual conventions as in Figure 2.

Acknowledgments

This work was supported by the Wellcome Trust Grant [206521/Z/17/Z] (K.M.J.), German Research Council (DFG) grants (CI 241/1-3, CI 241/1-7, INST 272/297-2) (R.M.C.), and

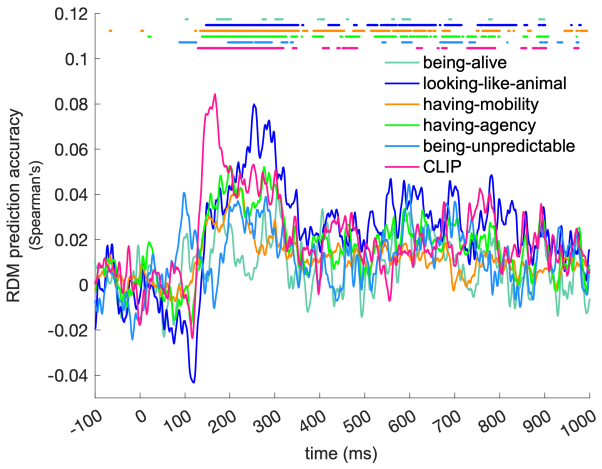

Figure 4: Animacy dimension RDM and CLIP comparison with EEG RDMs across time (lines - correlation between the EEG RDMs and each animacy dimension RDM, horizontal line above the graph - significant correlation (one-sided Wilcoxon signed-rank test, $p < 0.05$ corrected), grey horizontal bar on the x-axis - stimulus duration.

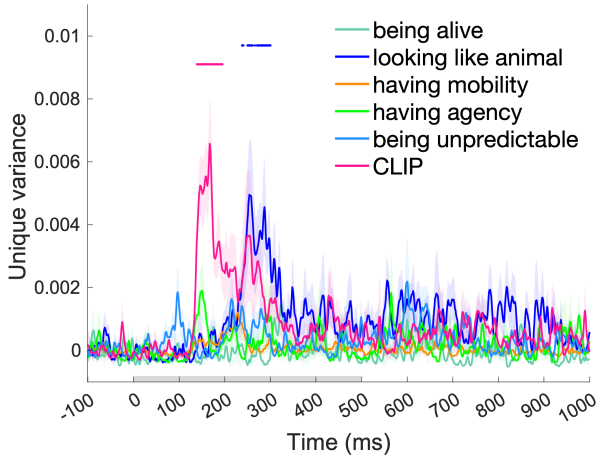

Figure 5: Unique variance of each animacy dimension and CLIP in explaining EEG RDMs computed using a GLM using the same visual conventions as in Figure 4.

European Research Council (ERC) Consolidator grant (ERC-CoG-2024101123101) (R.M.C.).

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., ... Kay, K. (2022). A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*, 116–126. doi: 10.1038/s41593-021-00962-x

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*, 149–159.

Bracci, S., Ritchie, J. B., Kalfas, I., & Op de Beeck, H. P. (2019). The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *Journal of Neuroscience*, *39*, 6513–6525.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*, 455–462. doi: 10.1038/nn.3635

Connolly, A. C., et al. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*, 2608–2618.

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., & Konkle, T. (2024). A large-scale examination of inductive biases shaping high-level visual representation in brains and machines. *Nature Communications*, *15*(9383). doi: 10.1038/s41467-024-53147-y

Hebart, M. N., Contier, O., Teichmann, L., Rockter, A. H., Zheng, C. Y., Kidder, A., ... Baker, C. I. (2023). Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife*, *12*, e82580. doi: 10.7554/eLife.82580

Jozwik, K. M., Kietzmann, T. C., Cichy, R. M., Kriegeskorte, N., & Mur, M. (2023). Deep neural networks and visuo-semantic models explain complementary components of human ventral-stream representational dynamics. *Journal of Neuroscience*, *43*(10), 1731–1741. doi: 10.1523/JNEUROSCI.1424-22.2022

Jozwik, K. M., Najarro, E., van den Bosch, J. J. F., Charest, I., Cichy, R. M., & Kriegeskorte, N. (2022). Disentangling five dimensions of animacy in human brain and behaviour. *Communications Biology*, *5*(1247). doi: 10.1038/s42003-022-04194-y

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. doi: 10.1016/j.neuron.2008.10.043

Lowder, M. W., & Gordon, P. C. (2015). Natural forces as agents: reconceptualizing the animate–inanimate distinction. *Cognition*, *136*, 85–90.

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-it object representation. *Frontiers in Psychology*, *4*, 128. doi: 10.3389/fpsyg.2013.00128

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of the 38th international conference on machine learning* (Vol. 139, pp. 8748–8763). PMLR.

Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human ventral temporal cortex. *eLife*, *8*, e47142.