Deep Neural Networks Provide Insights into Distinct and Shared Selectivity for Faces and Bodies in Human Visual Cortex

Leonard E. van Dyck (leonard.van-dyck@uni-giessen.de)
Department of Psychology
Justus Liebig University Giessen, Giessen, 35394, Germany
Max Planck Institute for Human Cognitive and Brain Sciences
Leipzig, 04103, Germany

Katharina Dobs (katharina.dobs@uni-giessen.de)
Department of Mathematics and Computer Science, Physics, Geography
Justus Liebig University Giessen, Giessen, 35392, Germany
Center for Mind, Brain and Behavior
Universities of Marburg, Giessen, and Darmstadt, Marburg, 35032, Germany

Abstract

Faces and bodies are key social stimuli processed by distinct functional networks in human visual cortex. However, growing evidence suggests systematic overlap between these networks, raising an important question: How segregated or integrated are face and body processing? Competing hypotheses propose fully segregated pathways, varying levels of integration, or a single multiplexed system. Here, we test these hypotheses using deep convolutional neural networks trained on object recognition. A functional localizer identified face- and body-selective units, as well as mixed-selective units that respond to both categories. Decoding analyses revealed that face- and body-selective units specialize in their respective domains, while mixed-selective units encode detailed information from both, supporting an integrative role. Finally, using fMRI encoding analyses, we found that these units account for unique variance in neural responses within both distinct and overlapping face- and body-selective cortical areas. Our findings suggest that face and body networks balance segregation and integration, supporting both fine-grained recognition and whole-person perception.

Keywords: functional selectivity; faces; bodies; deep neural networks; fMRI; visual cortex

Introduction

Faces and bodies are crucial for recognizing and understanding others. The human visual system is thought to process these stimuli in distinct functional networks, with specialized cortical areas dedicated to face and body perception (Freiwald et al., 2016; Peelen & Downing, 2007). However, growing evidence suggests systematic overlap between these networks (Schwarzlose et al., 2005), raising the question: How segregated or integrated are face and body processing? Competing hypotheses propose fully segregated pathways, varying levels of integration, or a single multiplexed system (Taubert et al., 2022). Deep neural networks (DNNs) offer a powerful framework to test these hypotheses and to explore, from a computational perspective, why face and body processing might be segregated or integrated (Kanwisher et al., 2023). While DNNs develop face- and body-selective units (Prince et al., 2024), it remains unclear whether they also form mixed-selective units that respond to both faces and bodies. Here, we examined whether a DNN trained on object recognition develops distinct face-, body-, and mixed-selective units, assessed their roles in domain-specific recognition, and evaluated their capacity to predict neural responses in both distinct and overlapping face- and body-selective areas.

Results

To investigate whether a DNN develops distinct mixed-selective units, we analyzed AlexNet (Krizhevsky et al., 2012) trained on Ecoset (Mehrer et al., 2021). Using a standard fMRI functional localizer (fLOC) with six stimulus categories (faces, bodies, scenes, characters, objects, and scrambled images; Stigliani et al., 2015), we identified face-, body-, and mixed-selective units (Fig. 1a). Specifically, in each layer, units were classified as face-selective if their activations (post-ReLU) to faces exceeded those to all other categories in pairwise *t*-tests (e.g., faces > bodies & faces > scenes, etc.; p<.05, FDR-corrected). Similarly, units were classified as body-selective if they preferred bodies over all other categories, and as mixed-selective if they preferred faces over all but bodies and vice versa. This analysis revealed distinct populations, with face-selective units being the most prevalent (max ~12%), and body- and mixed-selective units present in similar proportions (max ~2%; Fig. 1b). Selectivity of these units was quantified using d' based on the fLOC dataset and increased across layers (Fig. 1c). We validated this selectivity using an independent dataset containing faces, bodies, and objects (100 images each). Units retained their selectivity (Fig. 1d) but were distributed along a continuum of face-body selectivity (Fig. 1e), despite their initial discrete clustering, suggesting a graded rather than strictly categorical representational space.

Next, we examined the functional role of selective units in domain-specific processing by training linear classifiers (SVMs) to decode face identities and body parts from their activations. To ensure a fair comparison, we selected the same minimum number of top-ranked units per type in each layer. For face identity decoding, we used a dataset of 100 identities (10 images each; Dobs et al., 2023) with leave-one-exemplar-out cross-validation. Notably, face- and mixed-selective units outperformed body- and non-selective units in intermediate to late layers (Fig. 1f). For body part decoding, we used 6 categories from the THINGS database (e.g., arm, leg; 12-16 images each; Hebart et al., 2019) with stratified 5-fold cross-validation. Similarly, body- and mixed-selective units outperformed face- and non-selective units in intermediate layers (Fig. 1g). These findings suggest that mixed-selective units encode both face- and body-specific information, highlighting their potential role in integrating faces and bodies.

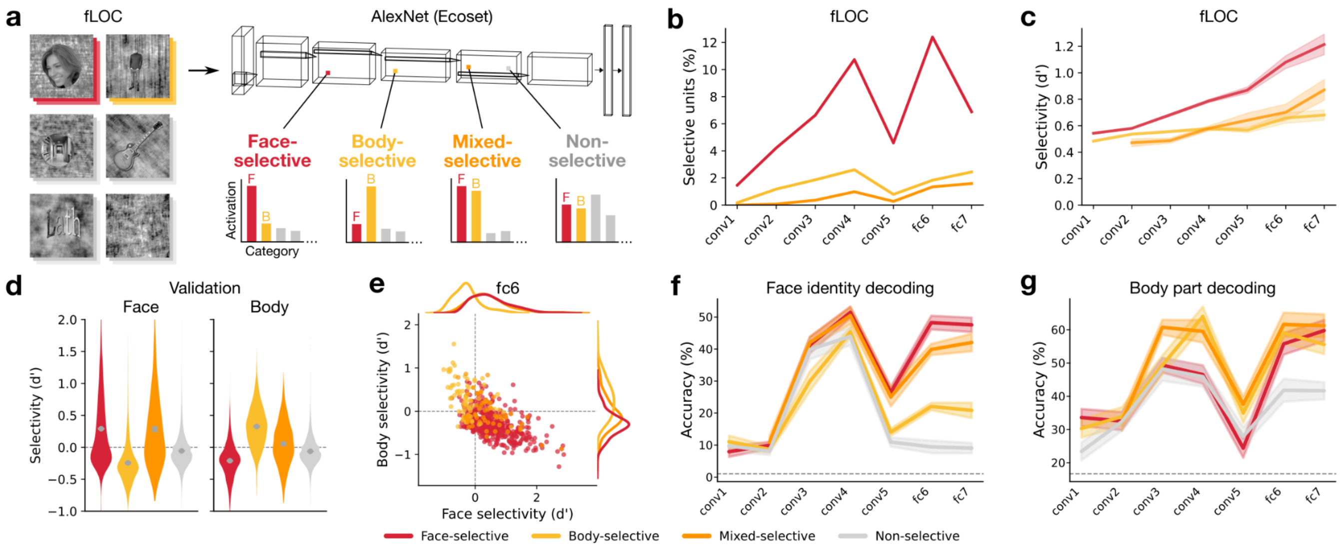

Figure 1: **a**, DNN fLOC approach. **b**, Proportion and **c**, selectivity (*d'*) of unit types across layers based on fLOC dataset. **d**, Layer-aggregated face and body selectivity of unit types, and **e**, individual unit selectivity in fc6 based on validation dataset. **f**, Face identity and **g**, body part decoding accuracy across layers. Shaded areas: 95% CIs.

Finally, we tested the correspondence between selective units and cortical areas using encoding models applied to fMRI responses from the Natural Scenes Dataset (Allen et al., 2022). We analyzed data from all participants (*N*=8), focusing on person-containing images (*n*~5000 per participant). Using the same fLOC (*t*>2; *SNR*>.15), we defined face- and body-selective areas (OFA, FFA, aTL-faces, EBA, FBA, mTL-bodies) and an overlapping area (FFA&FBA; Fig. 2a). To improve the interpretability of unit-to-voxel mappings, we predicted each voxel using sparse positive encoding via Lasso regression (*α*=.05, positivity constraint; Prince et al., 2024). This way, each voxel is explained by a purely additive combination of the most selective subset of units per type. A combined model with controlled unit counts per type was fit, and semi-partial correlation analysis isolated each type's unique contribution. Face- and body- selective units explained the most additional variance in their respective areas (Fig. 2b), confirming alignment with neural selectivity. Notably, mixed-selective units contributed comparably to both areas, suggesting shared coding mechanisms.

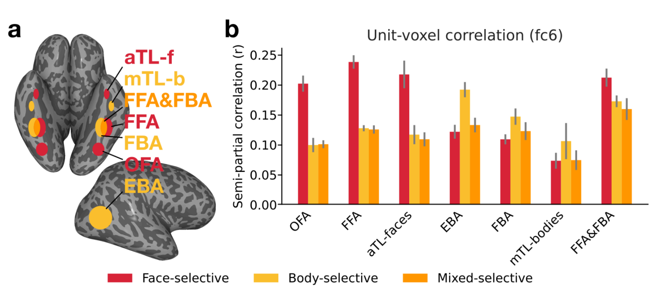

Figure 2: **a**, Schematic face- and body-selective areas. **b**, Semi-partial correlation of selective units in fc6 and voxels per area, averaged across participants. Error bars: 95% CIs.

Discussion

Our findings demonstrate that a DNN trained on object recognition develops distinct populations of face- and body-selective units, as well as mixed-selective units that respond to both. Unlike prior neuroimaging studies, where functional overlap may reflect lenient contrasts, our rigorous analyses reveal persistent entanglement. This pattern also generalized across architectures, training diets, and learning regimes (not displayed here). Mixed-selective units emerge early and become more prominent across processing stages, challenging notions of fully segregated or late-integrated pathways (Hu et al., 2020). Instead, they support progressively integrated networks (Harry et al., 2016) or a single multiplexed system with a face-body continuum (Tarhan & Konkle, 2020). Although mixed-selective units encode both face and body information and emerge in a task-optimized DNN in a way that suggests functional relevance, their precise role remains unclear. This motivates ongoing work exploring whether mixed-selective representations are essential for integrating face and body features into cohesive whole-person representations (Brandman & Yovel, 2016; Kaiser et al., 2014). In sum, our results reveal that DNNs exhibit both distinct and shared selectivity for faces and bodies, informing integration patterns in human visual cortex.

Acknowledgments

L.E.v.D was supported by a doctoral scholarship awarded by the German Academic Scholarship Foundation. K.D. was supported by the ERC Starting Grant DEEPFUNC (ERC-2023-STG-101117441), the Hessian Ministry of Higher Education, Research, Science and the Arts (LOEWE Start Professorship and Excellence Program "The Adaptive Mind"), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 222641018-SFB/TRR 135 TP C9).

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1). https://doi.org/10.1038/s41593-021-00962-x

Brandman, T., & Yovel, G. (2016). Bodies are Represented as Wholes Rather Than Their Sum of Parts in the Occipital-Temporal Cortex. *Cerebral Cortex*, *26*(2), 530–543. https://doi.org/10.1093/cercor/bhu205

Dobs, K., Yuan, J., Martinez, J., & Kanwisher, N. (2023). Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition. *Proceedings of the National Academy of Sciences*, *120*(32), e2220642120. https://doi.org/10.1073/pnas.2220642120

Freiwald, W., Duchaine, B., & Yovel, G. (2016). Face Processing Systems: From Neurons to Real-World Social Perception. *Annual Review of Neuroscience*, *39*, 325–346. https://doi.org/10.1146/annurev-neuro-070815-013934

Harry, B. B., Umla-Runge, K., Lawrence, A. D., Graham, K. S., & Downing, P. E. (2016). Evidence for Integrated Visual Face and Body Representations in the Anterior Temporal Lobes. *Journal of Cognitive Neuroscience*, *28*(8), 1178–1193. https://doi.org/10.1162/jocn_a_00966

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, *14*(10). https://doi.org/10.1371/journal.pone.0223792

Hu, Y., Baragchizadeh, A., & O'Toole, A. J. (2020). Integrating faces and bodies: Psychological and neural perspectives on whole person perception. *Neuroscience & Biobehavioral Reviews*, *112*, 472–486. https://doi.org/10.1016/j.neubiorev.2020.02.021

Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., & Peelen, M. V. (2014). Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *Journal of Neurophysiology*, *111*(1), 82–90. https://doi.org/10.1152/jn.00371.2013

Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask 'why' questions of minds and brains. *Trends in Neurosciences, 46*(3), 240-254. https://doi.org/10.1016/j.tins.2022.12.008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097–1105. https://doi.org/10.1145/3065386

Mehrer, J., Spoerer, C. J., Jones, E. C., Kriegeskorte, N., & Kietzmann, T. C. (2021). An ecologically motivated image dataset for deep learning yields better models of human vision. *Proceedings of the National Academy of Sciences*, *118*(8). https://doi.org/10.1073/pnas.2011417118

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*(8), 636–648. https://doi.org/10.1038/nrn2195

Prince, J. S., Alvarez, G. A., & Konkle, T. (2024). Contrastive learning explains the emergence and function of visual category-selective regions. *Science Advances*, *10*(39), eadl1776. https://doi.org/10.1126/sciadv.adl1776

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience*, *25*(47), 11055–11059. https://doi.org/10.1523/JNEUROSCI.2621-05.2005

Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*(1), 3002. https://doi.org/10.1038/s41467-020-16846-w

Taubert, J., Ritchie, J. B., Ungerleider, L. G., & Baker, C. I. (2022). One object, two networks? Assessing the relationship between the face and body-selective regions in the primate visual system. *Brain Structure and Function*, *227*(4), 1423–1438. https://doi.org/10.1007/s00429-021-02420-7