Modeling face-identity “likeness” with a convolutional neural network trained for face identification

Connor J. Parde[1,2], Alice J. O’Toole[1]

[1]The University of Texas at Dallas, [2]Vanderbilt University

Introduction and Goals

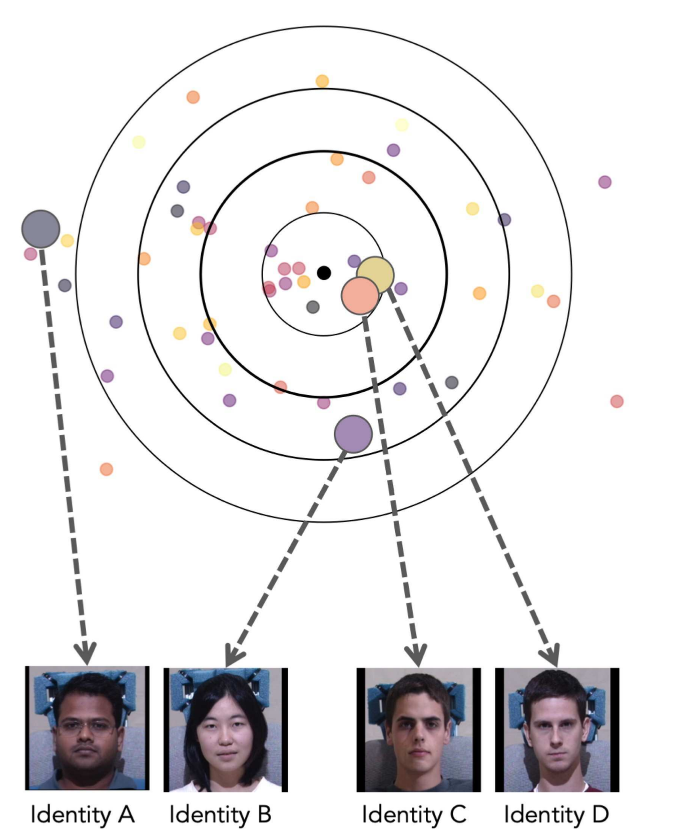
- Face identities vary in appearance (e.g., viewpoint, facial hair, expression, etc.)
- ***Perceived likeness***: *extent to which a face image is perceived to represent an identity accurately (‘good likeness’) or not (‘not a good likeness’)* (Ritchie et al., 2018)

- Inconsistent likeness ratings across participants (Hancock et al., 2009; White et al., 2017)
- Early ‘best likeness’: image averages (Brady et al., 2005), caricatures, and anti-caricatures (Lee et al., 2000; Kauffman et al., 2008)
- Recent best likeness: ‘iconic’ (Ritchie et al., 2018) or exemplar images (Balas et al., 2023)
- Mixed results due to lack of control over image parameters/observer experience
- Perceived likeness may not relate to similarity-to-prototype (Balas et al., 2023)

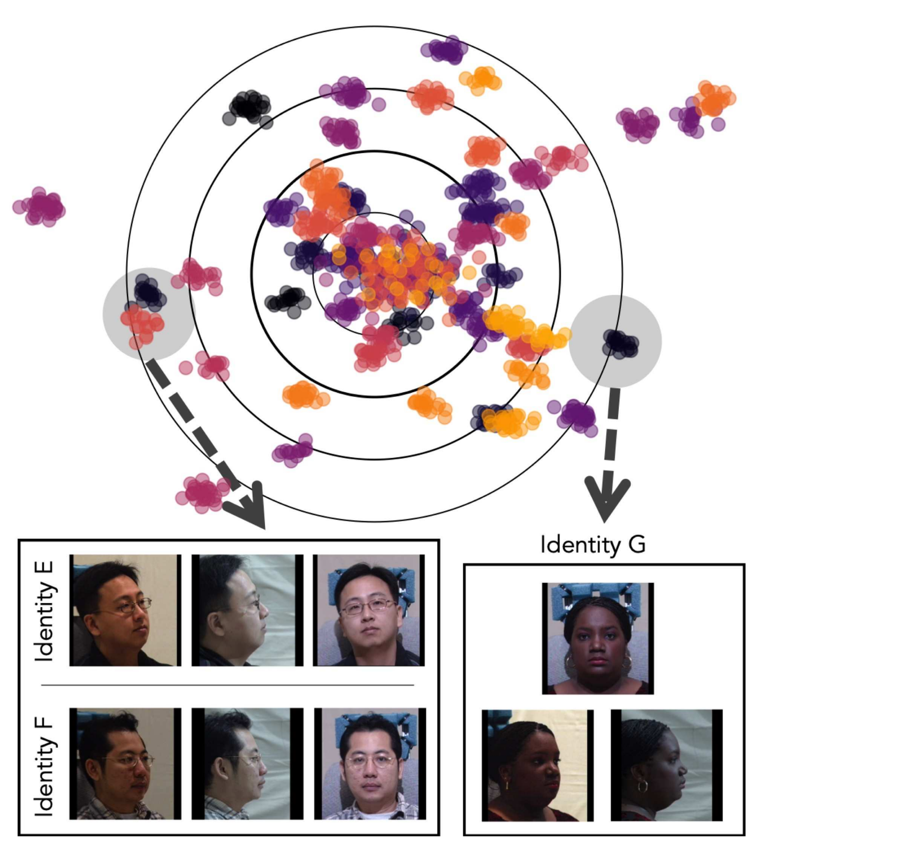
Deep Convolutional Neural Network (DCNN)-based Face Space

- Classical “face space” accounts for behavioral effects in face perception (e.g., inversion, ORE, caricaturing, etc.) *(Valentine, 1991; Valentine et al., 1992; Lee et al., 2000)*
- DCNN-based face space models within- and between-identity similarity *(Hill et al., 2019)*
- Natural testbed for modeling perceived likeness

Identity-based Face Space



DCNN-generated Face Space



Goal 1: Assess whether image parameters (i.e., viewpoint and illumination) by themselves affect the perceived likeness of a face image

Goal 2: Use a face space generated from a convolutional neural network to test competing models for measuring perceived likeness

Goal 3: Quantify how specific experience with an identity impacts perceived likeness of novel images showing the same identity

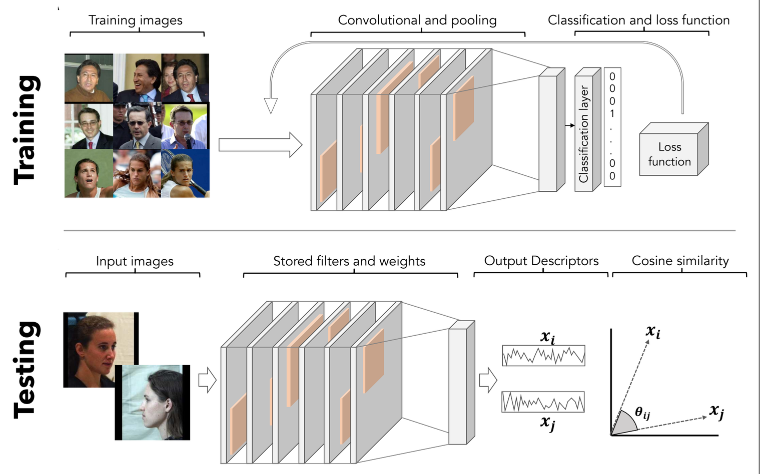
Dataset & DCNN

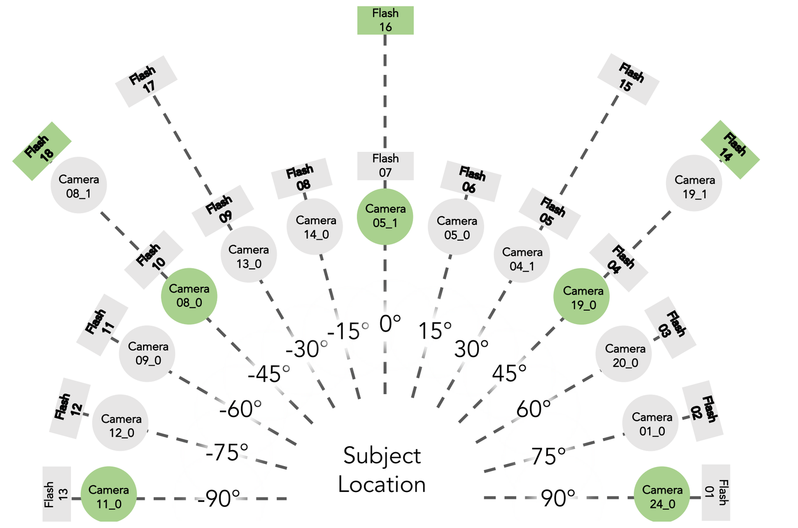
CMU Multi-PIE
(Gross et al., 2010)

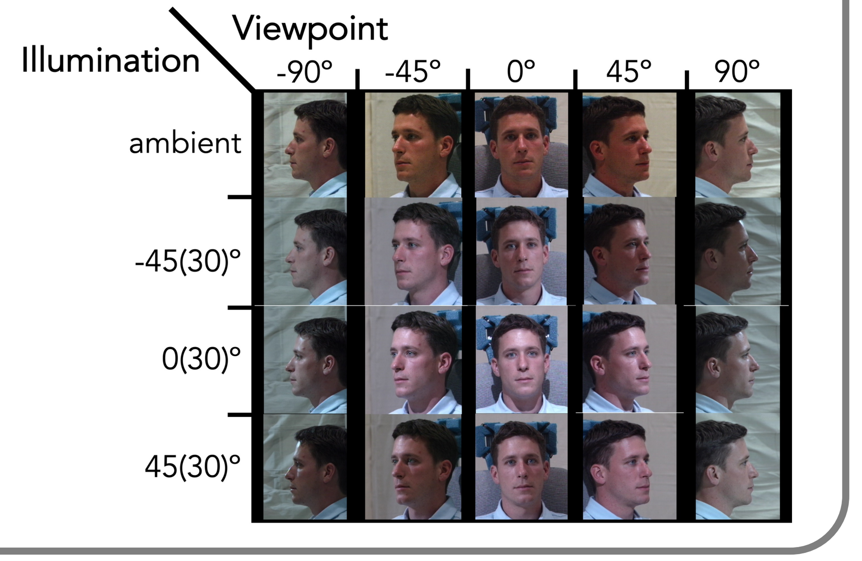
- Approx. 755k images of 337 non-celebrity identities
  - 60% Caucasian
- 129 subjects returned for 4 image capture sessions
- 15 viewpoints, 19 illumination conditions per session
- Images captured in rapid succession

Inception ResNet V1
(Sandberg, 2018)

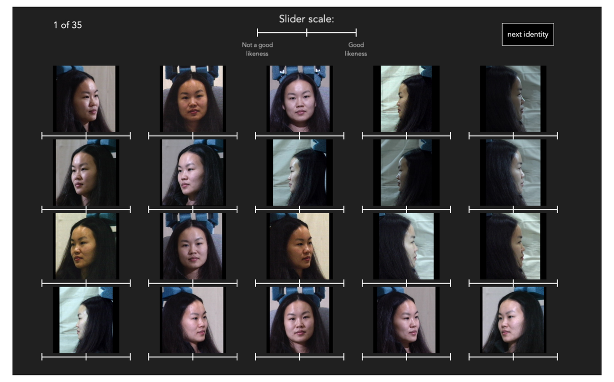
- “FaceNet” repository
- ResNet-101 architecture
- Pre-trained on VGGFace-2 (Cao et al., 2018)
- Face detection and alignment performed using MTCNN
- Extracted output from penultimate layer, 512-D descriptor vector per image







Experiment 1

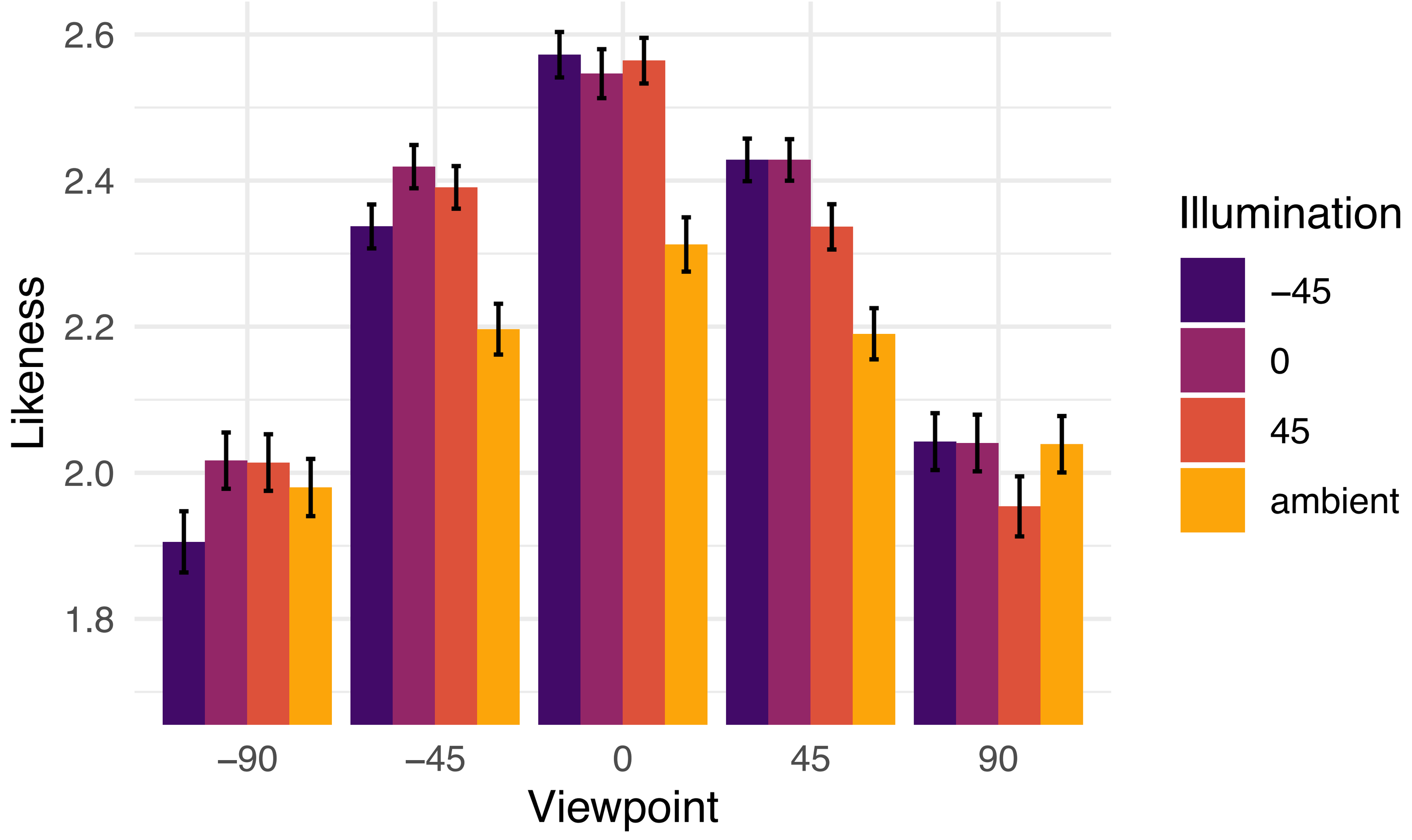


Are perceived-likeness ratings affected by viewpoint or illumination?

- 100 participants
- Controlled variation in image parameters across novel identities
- Simultaneously displayed all images of a given identity
- Participants adjusted slider bar to indicate whether each image was a ‘good likeness’ or ‘not a good likeness’
- Collapsed ratings across participants

- Submit averaged likeness ratings to 2-factor repeated measures ANOVA:
  - IV1 – viewpoint (-90°, -45°, 0°, 45°, 90°)
  - IV2 – illumination (ambient, -45°, 0°, 45°)
  - DV – perceived likeness rating

Results

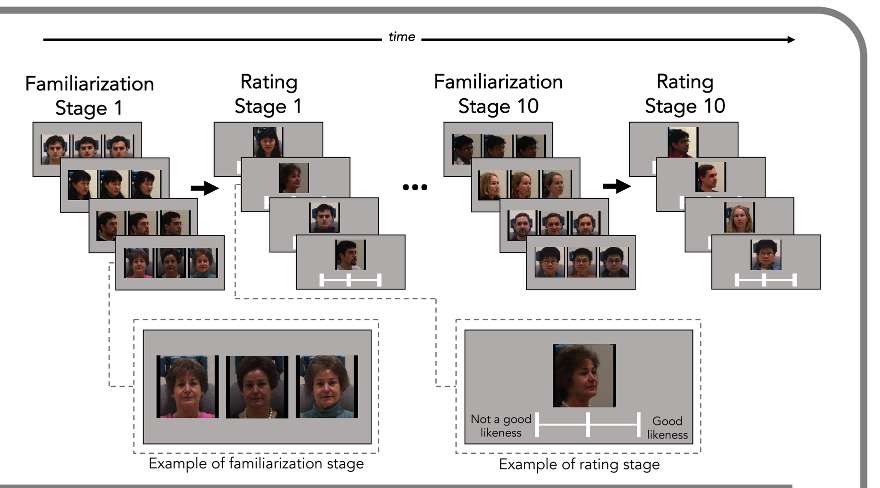


Viewpoint: ($F(4, 388) = 122.2$, $p < 0.0001$); Illumination: ($F(3, 291) = 62.07$, $p < 0.0001$); Interaction: ($F(12, 1164) = 57.03$, $p < 0$)

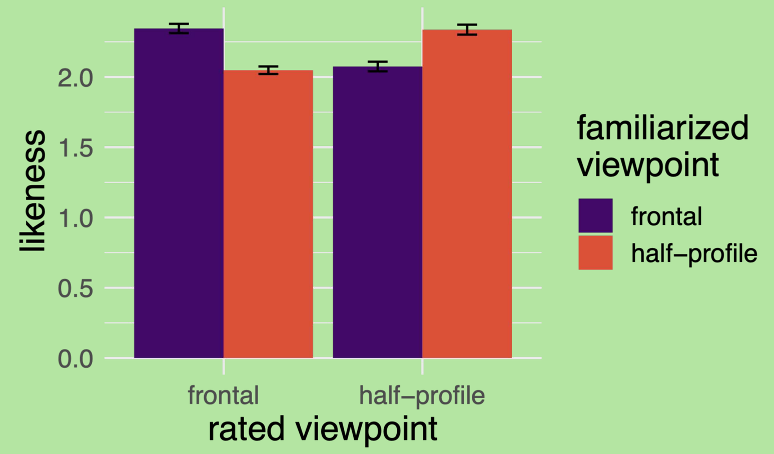
Experiment 2a (viewpoint) and 2b (illumination)

Do participants assign higher perceived-likeness ratings to face images that show an identity with the same viewpoint/illumination in which that identity was learned?

- Controlled viewing history with set of previously-unfamiliar identities
  - Identities learned with specific viewpoint (0° or 45°) or illumination (ambient or 45°)
- Collected likeness ratings for images with ***same*** or ***different*** viewpoint/illumination as seen previously
- Iterative stages of familiarization and rating (to accommodate working-memory load)
- Responses submitted to 2-factor ANOVA
  - IV1: familiarized viewpoint/illumination
  - IV2: rated viewpoint/illumination
  - DV: perceived likeness

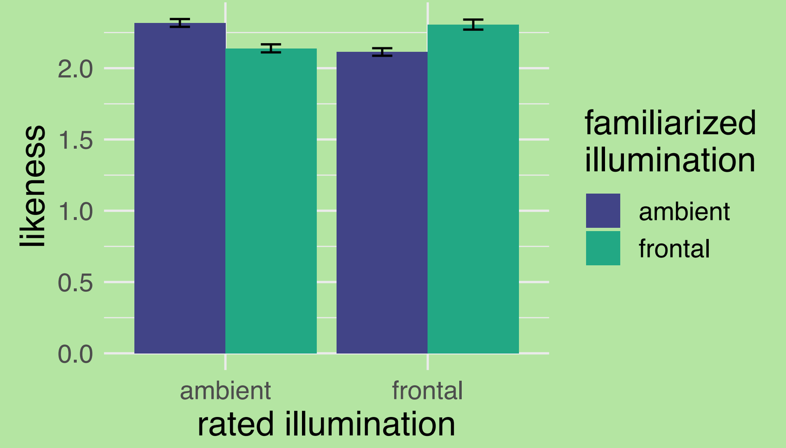


Viewpoint



Familiarized viewpoint: n.s.; Rated viewpoint: n.s.
Interaction: $F(1, 51) = 81.88$, $p < 0.0001$

Illumination

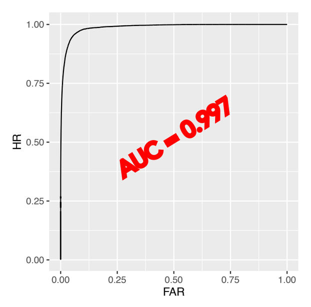


Familiarized illumination: n.s.; Rated illumination: n.s.
Interaction: $F(1, 51) = 75.17$, $p = < 0.0001$

Simulation 1

Can likeness be modeled using a DCNN?

- Same images from Experiment 1 processed through face-identification DCNN
- Generated likeness ratings relative to:
  1. *proximity to a central identity prototype*
  2. *local area density*
- Submit DCNN-based likeness ratings to 2-factor repeated measures ANOVA (as in Experiment 1)

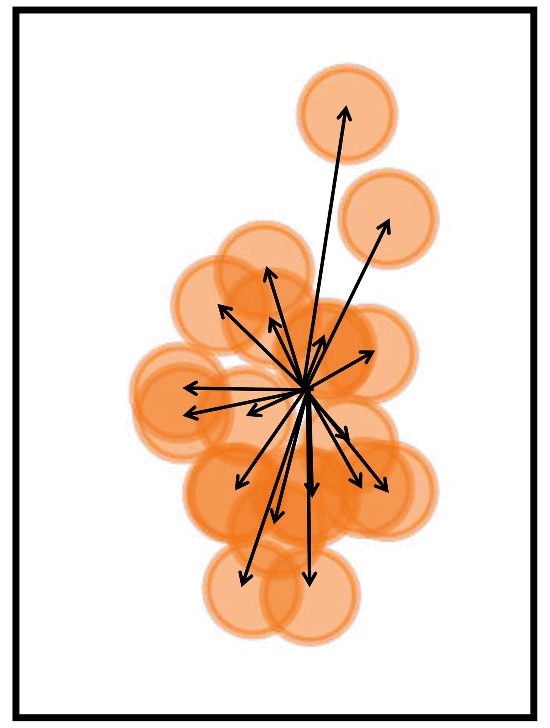


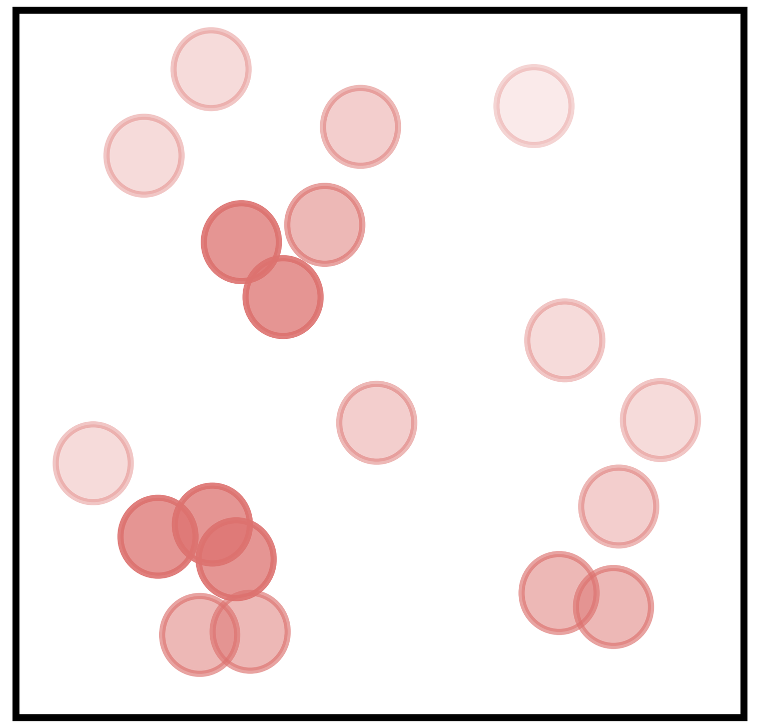
Proximity to Central Identity Prototype

- Average together descriptors generated for all images of each given identity
- Identity-specific averages defined as central identity prototype
- Measure distance of each image to respective central identity prototype
- Normalize distances for each identity

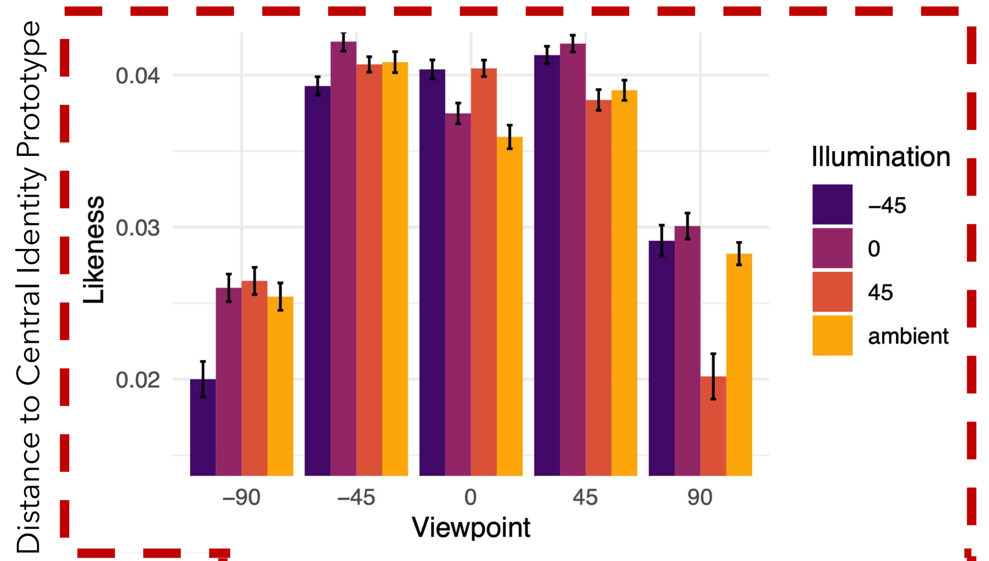
Local Area Density

- Generate distance matrix comparing cosine similarity of all pairs of same-identity descriptors
- Tally descriptors within one standard deviation of each point in identity-specific face space
- Number of “neighbors” indicates density of local area around each descriptor
- Descriptors with greater number of “neighbors” considered better likeness



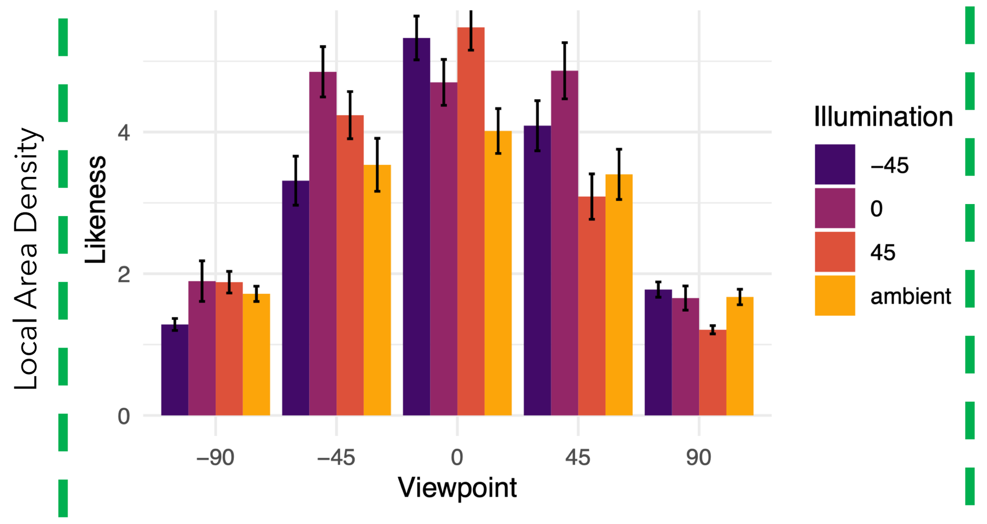


Results



Viewpoint: $F(4, 264) = 202.2$, $p < 0.0001$
Illumination: $F(3, 198) = 7.195$, $p = 0.00013$
Interaction: $F(12, 792) = 16.55$, $p < 0.0001$

Viewpoint: $F(4, 264) = 84.0$, $p < 0.0001$
Illumination: $F(3, 198) = 19.08$, $p < 0.0001$
Interaction: $F(12, 792) = 16.04$, $p < 0.0001$



Discussion:

Conclusion 1: In the absence of experience, viewpoint and illumination (and likely other cues) influence perceived-likeness.

Conclusion 2: Local-area density within an identity-specific, DCNN-based face space provides a more robust account for how human participants rate perceived likeness than proximity to a central identity prototype.

Conclusion 3: For previously-viewed identities, participants assign higher perceived-likeness ratings to images that match the viewpoint and illumination of previously-seen images of an identity.

Key Takeaways

Perceived likeness:

- Important to control within-identity variation when comparing perceived-likeness ratings across identities
- Similarity to identity prototypes only partially explain human ratings – distributed experience matters more
- Possible for visually distinct images to elicit similarly-high likeness ratings
  - Accounts for high likeness ratings assigned to both average-appearance images and anti-caricatures
- Experience matters - challenging to compare likeness ratings across raters without controlled experience

Modeling with CNNs:

- Utility of machine-learning models for testing psychologically-relevant hypotheses
- Given specific viewing experience, possible to estimate best perceived likeness of face identities

References

- Allen, H., N. Brady, and C. Tredoux (2009). Perception 38 (12), 1821–1830.
- Jenkins, R., D. White, X. Van Montfort, and A. M. Burton (2011). Cognition 121(3), 313–323.
- Ritchie, K. L., R. S. Kramer, and A. M. Burton (2018). Cognition 170, 1–8.
- Brady, N., M. Campbell, and M. Flaherty (2005). Brain and cognition 58 (3), 334–342.
- Lee, K., G. Byatt, and G. Rhodes (2000). Psychological science 11 (5), 379–385.
- Kaufmann, J. M. and S. R. Schweinberger (2008). Brain research 1228, 177–188.
- Hancock, J. T. and C. L. Toma (2009). Journal of Communication 59(2), 367–386.
- White, D., C. A. Sutherland, and A. L. Burton (2017). Cognitive research: principles and implications 2 (1), 1–9.
- Balas, B., Sandford, A., & Ritchie, K. (2023). i-Perception, 14(3).
- Valentine, T. (1991). The Quarterly Journal of Experimental Psychology Section A 43(2), 161–204.
- Valentine, T. and M. Endo (1992). The Quarterly Journal of Experimental Psychology 44 (4), 671– 703.
- Hill, M. Q., C. J. Parde, C. D. Castillo, Y. I. Colon, R. Ranjan, …, and A. J. O’Toole (2019). Nature Machine Intelligence 1(11), 522–529.
- Gross, R., I. Matthews, J. Cohn, T. Kanade, and S. Baker (2010). Image and vision computing 28(5), 807–813.
- Sandberg, D., FaceNet (2018). GitHub repository. https://github.com/davidsandberg/facenet.git
- Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018, May). 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 67-74). IEEE.

This work currently is accessible as a publicly-available dissertation document hosted by The University of Texas at Dallas