

The neural representational geometry of social perception

Jonathan B. Freeman, Ryan M. Stoler, Jeffrey A. Brooks, and Benjamin A. Stillerman

New York University

Corresponding author:

Jonathan B. Freeman
Department of Psychology
New York University
6 Washington Place
New York, NY 10003
Email: jon.freeman@nyu.edu

Abstract

An emerging focus on the geometry of representational structures is advancing a variety of areas in social perception, including social categorization, emotion perception, and trait impressions. Here, we review recent studies adopting a representational geometry approach, and argue that important advances in social perception can be gained by triangulating on the structure of representations via three levels of analysis: neuroimaging, real-time behavior (computer mouse-tracking), and neural-network modeling. This approach permits broad and comprehensive tests of how bottom-up facial features and visual processes as well as top-down social cognitive factors and conceptual processes shape perceptions of social categories, emotion, and personality traits. Although such work is only in its infancy, a focus on corroborating representational geometry across modalities is allowing researchers to use multiple levels of analysis to constrain theoretical models in social perception. This approach holds promise to further our understanding of the multiply determined nature of social perception and its neural basis.

THE NEURAL REPRESENTATIONAL GEOMETRY OF SOCIAL PERCEPTION

When we encounter other people, we instantly perceive the social categories to which they belong (e.g., gender, race), their current emotional state (e.g., sad), and even their personality traits (e.g., trustworthy, intelligent). Across all these instances of social perception lie underlying representations – of social categories, stereotypes, emotions, traits – that in turn drive perceptual judgments and behavior. Understanding social perception, surely, requires understanding these representations. Here, we propose that important advances in social perception can be gained by triangulating on the structure of such representations by relating three levels of analysis: neuroimaging, real-time behavior, and neural-network modeling. We will show how an emerging focus on the geometry of representational structures is advancing a variety of areas in social perception, including social categorization, emotion perception, and trait impressions.

Social perception as movement through multidimensional space

Neural-network models of social perception [1-4] assume that any given representation (e.g., male) is reflected by a unique pattern distributed over a population of nodes. It is the distributed pattern, dynamically re-instantiated in every new instance, that serves as the unique ‘code’ for a social category, stereotype, or trait. Such models are consistent with multi-cell recordings, which have shown that the communal activity of a population of neurons – a specific pattern of firing rates – provides the ‘code’ for various kinds of sensory and abstract cognitive information [i.e., a ‘population code’; 5].

We can conceive of representations in social perception (e.g., social categories, stereotypes, emotions, or traits) as points in a multidimensional space. Such a space can be measured using a variety of different modalities, such that the dimensions consist of neurons, fMRI voxels, or nodes in a neural-network model. The specific representational structure can be estimated through patterns of activation from these different modalities, as well as through seemingly unrelated measures (e.g. spatial dimensions in mouse-tracking or other behavioral data). Although these multidimensional spaces from different modalities may be radically different in an absolute sense, it is valuable to estimate the extent to which a shared representational geometry (i.e., the pairwise distances among representations) is preserved. This analytic approach – representational similarity analysis [RSA; 6,7] – can inform our

understanding not only of how representational spaces underlying social perception manifest at the neural level, but also help address fundamental psychological questions about how social perception is shaped by both relevant social cognitive and visual processes.

In certain neural-network models and in the actual brain, neural-representational patterns operate as ‘attractors’, such that a neural system is automatically attracted to complete those patterns when presented with a stimulus, allowing the system to descend from a high-energy state where the neuronal pattern is rapidly fluctuating to a lower-energy steady state in which the representational pattern comes to stabilize, i.e., an attractor state [8]. If the neuronal system’s state were imagined as a ball, the process of descending into an attractor state is analogous to a ball’s compulsion to roll down a hill, reflecting the increasing completion of the neural-representational pattern. Such attractors dynamics have long been observed in local populations of neurons in actual cortex [e.g., 9] and serve as an intrinsic pattern-completion process allowing neuronal patterns to serve as stored representations [e.g., 1,2].

Computational models. Imagine a system of 100 neurons, with two unique patterns of activation for the race categories White and Black; each pattern is therefore a point in a 100-dimensional space. The system, once presented with a person’s face, will move through 100-dimensional space toward the White or Black attractor state. Thus, at any moment, the neuronal population’s proximity to the White or Black attractor state in 100-dimensional space can be said to reflect to what extent that category representation is activated. Just as a ball’s energy is higher at the top of a hill relative to resting on the ground, each point in neural state space has an associated energy level, which determines the trajectory of where the system will go [10]. The system will gravitate toward energy minima, which are the stored representational patterns, i.e., attractor states [10]. If we were to project this 100-dimensional neural state space onto a more intuitive two-dimensional space, visualizing the energy levels at various points in the space, we can reveal these White and Black representations in the system (Fig. 1).

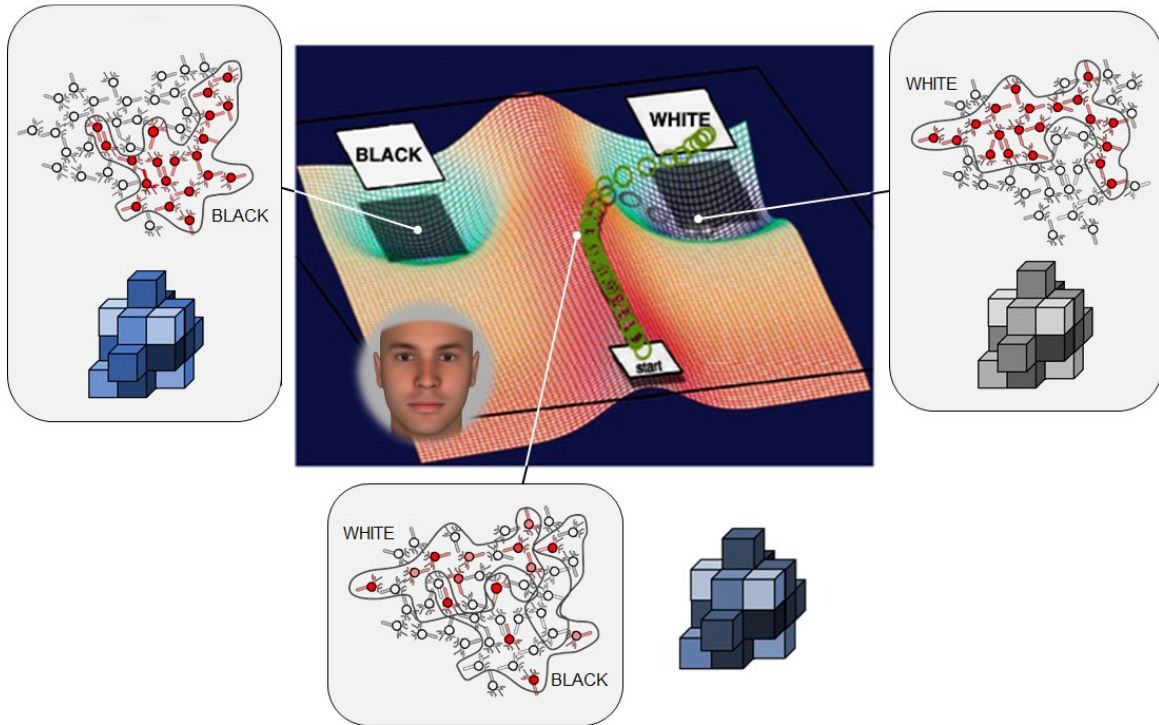


Figure 1. The link between modeling, mouse-tracking, and neuroimaging. An illustration of the link between the three modalities is depicted using an example of race categorization based on [11]. White and Black categories are each associated with a unique pattern across a neuronal population, with certain neurons highly active during a ‘White’ state and other neurons highly active during a ‘Black’ state. These states are low-energy attractors, into which the system is compelled to settle (similar to how a ball must roll down a hill). These states would be associated with distinct multi-voxel patterns using neuroimaging. Mouse-tracking can provide a real-time behavioral index of how the perceptual process settles over time into one of the two categories. The computer screen during mouse-tracking may serve as a two-dimensional proxy for higher dimensional neural state space. The mouse-tracking paradigm is depicted at the center, overlaid onto a hypothetical energy landscape describing the energy at all states in the system. The two energy minima (attractors) are shown, corresponding to the White and Black response locations and ‘White’ and ‘Black’ neural states (and corresponding multi-voxel patterns). At the beginning of the perceptual process, the system is in an unstable, high-energy state. As the process evolves over hundreds of milliseconds, the neuronal population gradually settles into a low-energy attractor state, i.e., White or Black category, just as the hand settles into one of the response locations. For a White-atypical face (with some Black-related features), during the perceptual process (e.g., mid-trajectory) the neuronal pattern would approximate the Black pattern to a greater extent and the hand would be more attracted toward the Black response. Because the multi-voxel pattern in response to such a face would reflect an average over this time period (as neuroimaging is not temporally sensitive), it would exhibit a degree of greater pattern-similarity to the Black category as well, as shown in previous neuroimaging work [11] (Box 1). This example shows the link between the multidimensional space of a neuronal population/model, the two-dimensional space of a computer screen with mouse-tracking, and the multidimensional space of multi-voxel response patterns.

Mouse-tracking. To measure these attractor dynamics during social perception, researchers have recently leveraged response-directed hand movement using mouse-tracking [12] At any given moment of time in a mouse-tracking task, the hand’s proximity to a given

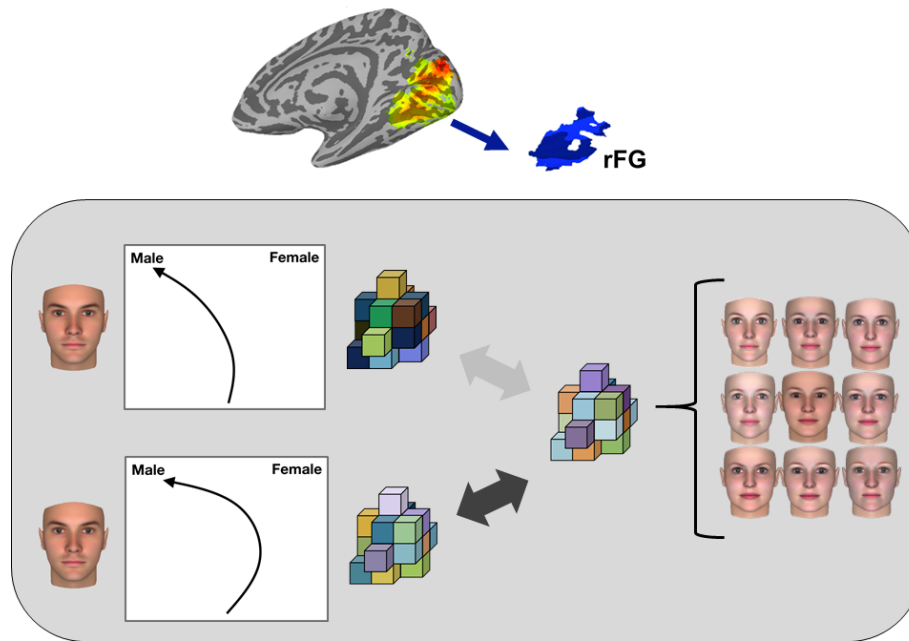
response (e.g., a social category, emotion, or trait) on the two-dimensional space of the screen – like the proximity of the system to a given attractor in high-dimensional neural state space – may index the extent to which that social category, emotion, or trait representation is currently activated. By examining how the hand settles into a response over time, and may be partially pulled toward other potential responses, mouse-tracking has charted out the real-time dynamics through which social categories, emotions, stereotypes, attitudes, and traits activate and resolve over hundreds of milliseconds [12,13]. The distance the hand travels toward an unselected response option (before arriving at the selected option) can be taken as a proxy – using the physical distance in two-dimensional space – of the distance between those two representations in higher-dimensional neural state space (Fig. 1).

Neuroimaging. Finally, computational models and mouse-tracking can be combined with neuroimaging. Neuroimaging studies have increasingly focused on multi-voxel patterns of activity within a functional region to understand representational structure [6,7]. Although a far coarser measure than measuring activity of actual neurons, individual voxels may contain different distributions of neurons selective for certain types of information, thereby allowing voxel patterns to serve as potential proxies of underlying neuronal pattern [14,15]. Recent multi-voxel pattern analyses have demonstrated reliable decoding of various kinds of representations in social perception. For instance, studies have shown the fusiform gyrus (FG), a visual face-processing region [16], is involved in representing faces' gender [17], race [18,19], and emotion [20]. Facial emotion categories have also been decoded from the superior temporal sulcus, a region important for processing dynamic visual cues [16,21]. Consider a brain region with 100 voxels, with two unique multi-voxel patterns associated with the race categories White and Black; each category representation is therefore a point in a 100-dimensional multi-voxel space. Just like distances among representations in the multidimensional space of a neural network, or in the two-dimensional space of a computer screen with mouse-tracking, distances in the multidimensional space of a brain region's response patterns can be used to understand the structure of representations in social perception (Fig. 1).

Linking these levels of analysis – computational modeling, mouse-tracking, and neuroimaging – by comparing their relative representational geometries has recently been used to address theoretically important questions in social perception.

Social categorization

A glimpse of a face conveys a number of social categories, such as gender or race. Such categorizations are consequential, as they tend to automatically activate related stereotypes and attitudes and impact behavior in unintended ways [22,23]. A central challenge of social categorization is to take the inherent diversity in facial features out in the social world to form a coherent categorization. Current neural-network models predict that when facial features occasionally conflict (e.g., a man's face with feminine features), as with natural variation in the population, multiple partially-active social categories will be activated (i.e., both male and female) that simultaneously compete over time [2,25] – a process for which representational geometry has recently been useful in understanding [11] (Box 1).



Box 1. Co-activation of representations in social categorization. A recent study [11] synchronizing neuroimaging and mouse-tracking found that, when categorizing gender- or race-atypical faces (e.g., a feminine male face), participants' hand trajectories exhibited a parallel attraction toward the opposite gender or race category response before ultimately arriving at the correct response, showing that the opposite category was co-activated in parallel. In the face-processing right FG, each participant's distinct multi-voxel pattern for the male, female, White, and Black category was assessed. The results showed that the extent of parallel attraction toward the opposite category (e.g., toward 'female' for a male face) was associated with a greater neural-pattern similarity to that opposite category's multi-voxel pattern (or toward 'Black' for a White face with Black-related features, see Fig. 1). Such work suggests that common ambiguities tend to activate alternate social categories before a categorization stabilizes, and this is reflected in the similarity of that face's multi-voxel pattern to the alternate social category's pattern in the right FG [11]. Thus, by examining the geometry of a given face's neural response pattern to the patterns associated with other social categories, novel insights may be made into the mechanisms of social categorization.

This category competition driving perception may also be affected by top-down factors as well. For instance, merely being exposed to a face has long been known to spontaneously activate relevant gender or racial stereotypes [22,24], but recent work has suggested a more bidirectional relationship between face perception and stereotypes. Like other forms of perception, top-down expectations may facilitate or inhibit certain perceptual interpretations [25,26], which in social perception may include stereotype-based expectations [2,23]. Over the years, stereotype effects have been documented on perceptual judgments [27-30], but it has remained less clear how “deeply” such biases manifest.

In one set of neuroimaging studies [31], the structure (i.e., all pairwise similarities) of gender, race, and emotion categories was assessed across several domains: multi-voxel response patterns to faces (via neuroimaging), subjective perception of faces (via mouse-tracking), stereotype knowledge (via a stereotype content task), and faces’ intrinsic visual properties (via visual stimulus models). The results revealed a shared representational geometry across modalities (stereotype knowledge, subjective perceptions, and multi-voxel response patterns in the face-processing right FG). The similarity (i.e., distance) between any given pair of social categories (e.g., Black and anger) in stereotype knowledge predicted a corresponding similarity in how faces belonging to those categories were perceived (via mouse-tracking) as well as in multi-voxel response patterns to those faces. For example, the more similarly the categories ‘Black’ and ‘anger’ were believed to be in terms of their social-conceptual knowledge (i.e. stereotypes) predicted a greater tendency to perceive Black faces and angry faces more similarly (e.g., a partial attraction to ‘angry’ even for a non-angry Black face). This greater perceptual similarity was in turn reflected by an increased similarity in the multi-voxel representations of Black faces and angry faces in the FG (Fig. 2a). Importantly, the relationship between stereotype structure, perceptual structure, and neural-pattern structure held even when statistically controlling for stimuli’s intrinsic physical structure, thereby ensuring the effects reflected stereotypes rather than visual resemblances in the stimuli (e.g., between Black and angry faces). That these stereotypical biases on perception manifest in a visual face-processing region’s representational structure suggests that social cognitive processes can impact relatively low levels of visual representation.

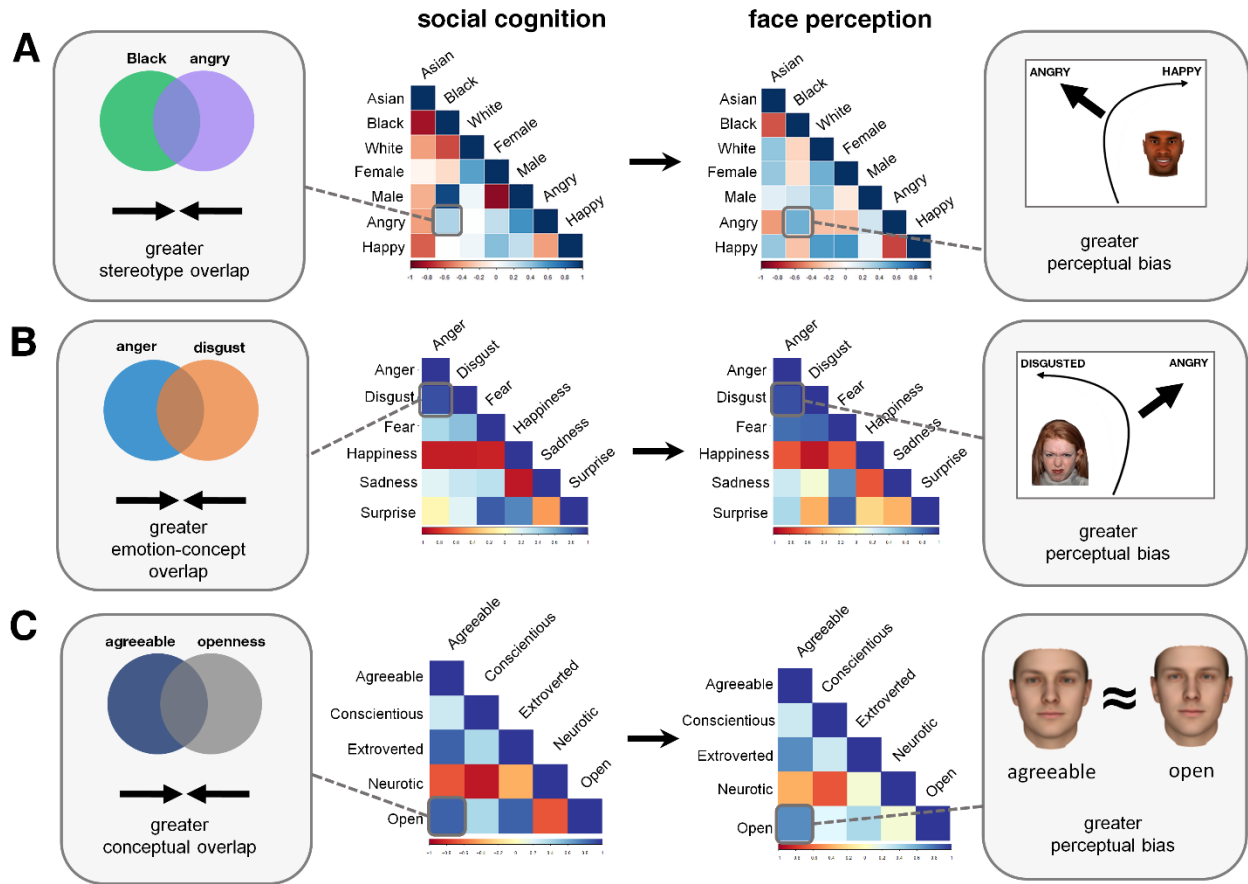


Figure 2. Representational similarity analysis (RSA) in social perception. In RSA, representational dissimilarity matrices (RDMs) comprise all pairwise similarities/dissimilarities and are estimated for each modality. Unique values under the diagonal are vectorized, with each vector reflecting the geometry of the representational space. Correspondence in geometry is then assessed through the vectors' bivariate relationships (e.g., correlation, regression). **(A)** Participants' stereotype RDM (stereotype content task) predicted their perceptual RDM (mouse-tracking), showing that a biased similarity between two social categories in stereotype knowledge was associated with a bias to see faces belonging to those categories more similarly, which in turn was reflected in FG pattern structure [31]. **(B)** Participants' emotion-concept RDM (emotion ratings task) predicted their perceptual RDM (mouse-tracking), showing that an increased similarity between two emotion categories in emotion-concept knowledge was associated with a tendency to perceive those facial expressions more similarly [32], which was also reflected in FG pattern structure [33]. **(C)** Participants' conceptual RDM (trait ratings task) predicted their perceptual RDM (reverse correlation task), showing that an increased tendency to believe two traits are conceptually more similar is associated with using more similar facial features to make inferences about those traits [34].

Emotion perception

As in social categorization, the notion that cognitive processes may exert top-down impacts on perception has become increasingly important in understanding emotion perception, particularly the role of emotion-concept knowledge [35,36]. Early basic-emotion approaches

emphasized six universal facial expressions of emotion – anger, joy, sadness, disgust, fear, and surprise – that are associated with specific action units that may be read directly from a face. Constructionist approaches and recent neural-network models suggest that not only do facial features related to, fear, for example, drive categorizations, but so does conceptual knowledge about what fear means. To examine this conceptual influence, previous research manipulated particular emotion concepts or measured emotion perception in patients with semantic dementia, who have reduced emotion-concept access [37,38], but a comprehensive test of how inter-individual variability in conceptual knowledge may shape facial emotion perception across the broad spectrum of emotion categories has been lacking.

In one set of studies [32], the structure (i.e., pairwise similarities) of the six basic emotion categories was assessed at the level of conceptual knowledge (via a rating task), facial emotion perception (via mouse-tracking) and faces' intrinsic physical properties (via measuring facial action units). A corresponding geometry was observed between conceptual knowledge and facial emotion perception. When individuals believed two emotions (e.g., anger and disgust) to be conceptually more similar, faces belonging to those categories were perceived with a corresponding similarity (i.e., mouse trajectories were more attracted to both emotion responses in parallel, although each face only depicted one emotion), even when controlling for any possible physical similarity between the two kinds of faces (Fig. 2b). Such findings suggest that subtle individual differences in the conceptual understanding of what different emotions mean are reflected in how those emotions are perceived from a face. A neuroimaging study [33] replicated this general finding but also showed that such a conceptually shaped representational structure was reflected in the structure of FG multi-voxel patterns, thereby demonstrating its impact on relatively low-level visual representation. Thus, believing anger and disgust are more conceptually related is associated with an increased bias to perceive angry and disgusted faces more similarly, as well as increased similarity in their multi-voxel response patterns in the FG.

Other recent work has used representational geometry to test competing models of inferring the emotional experience of another person [39]. In response to emotional episodes, the similarity structure of multi-voxel patterns in the temporoparietal junction and dorsomedial prefrontal cortex, regions implicated in theory of mind, were compared to the similarity structure of several candidate theoretical models. Among these was an appraisal model, relating episodes by their similarity across 38 appraisal dimensions (i.e., abstract features of the causal contexts

that elicit emotions). The geometry of neural patterns in theory of mind regions corresponded best with the geometry of this 38-dimensional appraisal model, suggesting that these regions contain a high-dimensional model capturing the appraisal of others' emotional experiences that cannot be reduced merely to a two-dimensional circumplex model (valence and arousal) or six-dimensional basic-emotion model.

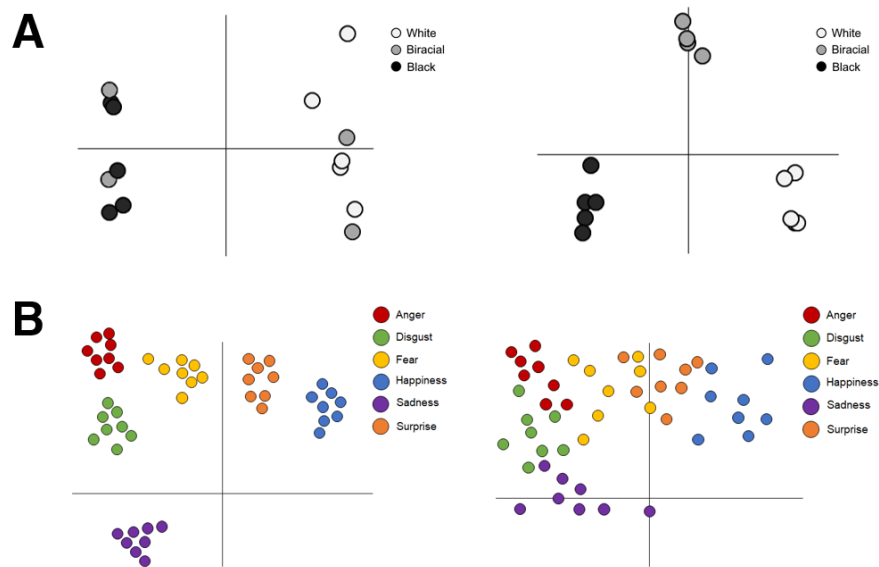
Trait impressions

Perceiving personality traits from a face – warmth, intelligence – may seem ambiguous and arbitrary relative to perceiving aspects like social categories or emotions. But much research has demonstrated that arrangements of facial features reliably relate to specific trait perceptions [40-42], a process requiring minimal visual exposure [43,44]. Popular models of face-based trait impressions focus exclusively on the role of bottom-up facial features [40]. However, an emerging literature has noted the considerable influence of top-down social cognitive processes in face impressions [41,45]. Advancing these perspectives, recent work has used representational geometry to examine how social cognitive processes or conceptual knowledge may shape the structure of forming trait impressions from a face.

A recent set of studies [34] tested whether individual differences in a perceiver's conceptual trait associations shape how those traits are perceived from a face. The structure (i.e., all pairwise similarities) of traits was assessed at the level of conceptual knowledge (e.g., “if someone is agreeable, are they also open-minded?”) and of face impressions. A reverse correlation technique was used to estimate each perceiver's visual prototypes for different traits. Indeed, to the extent a perceiver believed any pair of traits to be more related (e.g., openness and agreeableness), they perceived those traits in faces with a corresponding similarity (e.g., greater resemblance in the visual prototypes for the two traits; Fig. 2c). This showed that face impressions arise not just from bottom-up features but are also scaffolded by the conceptual understanding of what those traits mean. This opens the door to inter-individual variability and departs from prior models of face impressions that focus on a relatively fixed and universal architecture.

Other recent research went so far as to quantify the role of conceptual trait knowledge in social perception generally – across domains of face impressions, familiar person knowledge, and group stereotypes [46]. Analyses demonstrated a strong correspondence in the geometry of

trait representations across these social perception models, suggesting that a perceiver's conceptual trait associations may provide a domain-general model for inferring about not only faces but also familiar others and social groups. Finally, representational geometry has additionally been used to understand inferences about others' transient mental states, finding that a three-dimensional model of rationality, social impact, and valence best predict the structure of behavioral data and neural patterns associated with mental states [47].



Box 2. Representational geometries. Besides testing correspondence between representational geometries of different modalities using RSA (Fig. 2), their clustering and organization can also be informative. Hypothetical geometries (from multi-voxel patterns, behavioral data, or model simulations) are provided in a reduced, intuitive two-dimensional space. **(A)** When presented with White, Biracial, and Black faces, one brain region's multi-voxel patterns may have a two-category organization, placing Biracial faces into either the White or Black cluster (left) whereas another region may have a three-category organization, such that Biracial faces are placed into their own distinct cluster (right). **(B)** When presented with emotional faces, certain perceivers may have a six-category organization in behavioral data or neural patterns, such that each basic emotion has a distinct cluster (left), whereas other perceivers may have a more blended organization, such that various emotion expressions do not fit into the six distinct emotion categories [e.g., 48]. Thus, examining representational spaces in social perception may reveal important differences in the perceptual organization of different brain regions or individual perceivers.

Conclusion

Across various domains of social perception, an emerging focus on representational geometry is permitting broad and comprehensive tests of how both bottom-up facial features and a variety of top-down social cognitive factors together shape perceptions of other people. Linking across neural-network modeling, neuroimaging, and behavioral techniques such as

mouse-tracking is allowing researchers to use multiple levels of analysis to constrain theoretical models. Although such work is in its early stages, it provides promise for furthering our understanding of the multiply determined nature of social perception and its neural basis.

Acknowledgements

This work was supported in part by research grants NSF BCS-1654731 and NIH R01-MH112640 to J.B.F.

References

1. Smith ER, DeCoster J: **Knowledge acquisition, accessibility, and use in person perception and stereotyping: Simulation with a recurrent connectionist network.** *Journal of Personality and Social Psychology* 1998, **74**:21-35.
2. Freeman JB, Ambady N: **A dynamic interactive theory of person construal.** *Psychological Review* 2011, **118**:247-279.
3. Read SJ, Miller LC: **On the dynamic construction of meaning: An interactive activation and competition model of social perception.** In *Connectionist models of social reasoning and social behavior*. Edited by Read SJ, Miller LC: Erlbaum; 1998.
4. Kunda Z, Thagard P: **Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory.** *Psychological Review* 1996, **103**:284-308.
5. Auerbeck BB, Latham PE, Pouget A: **Neural correlations, population coding and computation.** *Nature reviews neuroscience* 2006, **7**:358.
6. Kriegeskorte N, Kievit RA: **Representational geometry: integrating cognition, computation, and the brain.** *Trends in cognitive sciences* 2013, **17**:401-412.
7. Kriegeskorte N, Mur M, Bandettini PA: **Representational similarity analysis-connecting the branches of systems neuroscience.** *Frontiers in systems neuroscience* 2008, **2**:4.
8. Smolensky P: **Connectionist modeling: Neural computation/mental connections.** In *Neural connections, mental computations*. Edited by Nadel L, Cooper A, Culicover P, Harnish RM: MIT Press; 1989.
9. Cossart R, Aronov D, Yuste R: **Attractor dynamics of network UP states in the neocortex.** *Nature* 2003, **423**:283.
10. Hopfield JJ: **Neural networks and physical systems with emergent collective computational abilities.** *Proceedings of the National Academy of Sciences of the United States of America* 1982, **79**:2554-2558.
11. Stolier RM, Freeman JB: **A neural mechanism of social categorization.** *Journal of Neuroscience* 2017, **37**:5711-5721.
12. Freeman JB: **Doing psychological science by hand.** *Current Directions in Psychological Science* 2018.
13. Freeman JB, Dale R, Farmer TA: **Hand in motion reveals mind in motion.** *Frontiers in Psychology* 2011, **2**:59.

14. Chaimow D, Yacoub E, Ugurbil K, Shmuel A: **Modeling and analysis of mechanisms underlying fMRI-based decoding of information conveyed in cortical columns.** *Neuroimage* 2011, **56**:627-642.
15. Haynes J-D: **A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives.** *Neuron* 2015, **87**:257-270.
16. Haxby JV, Hoffman EA, Gobbini MI: **The distributed human neural system for face perception.** *Trends in Cognitive Sciences* 2000, **4**:223-233.
17. Freeman JB, Rule NO, Adams RB, Ambady N: **The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices.** *Cerebral Cortex* 2010, **20**:1314-1322.
18. Contreras JM, Banaji MR, Mitchell JP: **Multivoxel patterns in fusiform face area differentiate faces by sex and race.** *PloS one* 2013, **8**:e69684.
19. Ratner KG, Kaul C, Van Bavel JJ: **Is race erased? Decoding race from patterns of neural activity when skin color is not diagnostic of group boundaries.** *Social cognitive and affective neuroscience* 2013, **8**:750-755.
20. Wegrzyn M, Riehle M, Labudda K, Woermann F, Baumgartner F, Pollmann S, Bien CG, Kissler J: **Investigating the brain basis of facial expression perception using multi-voxel pattern analysis.** *Cortex* 2015.
21. Winston JS, Henson RNA, Fine-Goulden MR, Dolan RJ: **fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception.** *Journal of Neurophysiology* 2004, **92**:1830-1839.
22. Macrae CN, Bodenhausen GV: **Social cognition: Thinking categorically about others.** *Annual Review of Psychology* 2000, **51**:93-120.
23. Freeman JB, Johnson KL: **More than meets the eye: Split-second social perception.** *Trends in Cognitive Sciences* 2016.
- ** **Review of visually based social perception and its neural basis, including how mouse-tracking, multi-voxel patterns, modeling, and attractor dynamics can be used to understand the process of social perception.**
24. Mason MF, Cloutier J, Macrae CN: **On construing others: Category and stereotype activation from facial cues.** *Social Cognition* 2006, **24**:540-562.
25. Bar M: **Visual objects in context.** *Nature Reviews Neuroscience* 2004, **5**:617-629.

26. Summerfield C, Egner T: **Expectation (and attention) in visual cognition.** *Trends in cognitive sciences* 2009, **13**:403-409.
27. Hugenberg K, Bodenhausen GV: **Ambiguity in Social Categorization: The role of prejudice and facial affect in race categorization.** *Psychological Science* 2004, **15**:342-345.
28. Hugenberg K, Bodenhausen GV: **Facing prejudice: Implicit prejudice and the perception of facial threat.** *Psychological Science* 2003, **14**:640-643.
29. Johnson KL, Freeman JB, Pauker K: **Race is gendered: How Covarying Phenotypes and Stereotypes Bias Sex Categorization.** *Journal of Personality and Social Psychology* 2012:doi: 10.1037/a0025335.
30. Hess U, Senécal S, Kirouac G, Herrera P, Philippot P, Kleck RE: **Emotional expressivity in men and women: Stereotypes and self-perceptions.** *Cognition & Emotion*, **14,5** 2000.
31. Stoller RM, Freeman JB: **Neural pattern similarity reveals the inherent intersection of social categories.** *Nature Neuroscience* 2016, **19**:795-797.
- * **A set of neuroimaging studies using representational geometry, finding that stereotypes shape perceptions of faces' social categories.**
32. Brooks JA, Freeman JB: **Conceptual knowledge predicts the representational structure of facial emotion perception.** *Nature Human Behaviour* in press.
- * **A set of studies using representational geometry, finding that emotion-concept knowledge shapes perceptions of faces' emotional expressions.**
33. Brooks JA, Chikazoe J, Sadato N, Freeman JB: **The neural representation of emotion perception reflects cultural and individual variability in conceptual knowledge.** in prep.
34. Stoller RM, Hehman E, Keller M, Walker M, Freeman JB: **The conceptual structure of face impressions.** *PNAS* invited revision.
35. Barrett LF: *How emotions are made: The secret life of the brain*: Houghton Mifflin Harcourt; 2017.
36. Lindquist KA, Barrett LF: **Constructing Emotion The Experience of Fear as a Conceptual Act.** *Psychological science* 2008, **19**:898-903.
37. Gendron M, Lindquist KA, Barsalou L, Barrett LF: **Emotion words shape emotion percepts.** *Emotion* 2012, **12**:314-325.

38. Lindquist KA, Gendron M, Barrett LF, Dickerson BC: **Emotion perception, but not affect perception, is impaired with semantic memory loss.** *Emotion* 2014, **14**:375-387.
39. Skerry AE, Saxe R: **A common neural code for perceived and inferred emotion.** *Journal of Neuroscience* 2014, **34**:15997-16008.
- *A neuroimaging study using representational geometry to demonstrate that theory of mind regions best reflect a high-dimensional appraisal model of others' emotional experience, rather than tracking merely valence and arousal or the six basic emotions.
40. Oosterhof NN, Todorov A: **The functional basis of face evaluation.** *Proceedings of the National Academy of Sciences* 2008, **105**:11087-11092.
- *A foundational model of trait impressions of faces, focusing on how specific sets of facial features relate to particular trait inferences.
41. Hehman E, Sutherland CA, Flake JK, Slepian ML: **The unique contributions of perceiver and target characteristics in person perception.** *Journal of personality and social psychology* 2017, **113**:513.
42. Zebrowitz LA, Montepare JM: **Social psychological face perception: Why appearance matters.** *Social and Personality Psychology Compass* 2008, **2**:1497-1517.
43. Willis J, Todorov A: **First impressions: Making up your mind after a 100-ms exposure to a face.** *Psychological Science* 2006, **17**:592-598.
44. Freeman JB, Stoler RM, Ingbretsen ZA, Hehman EA: **Amygdala responsivity to high-level social information from unseen faces.** *The Journal of Neuroscience* 2014, **34**:10573-10581.
45. Stoler RM, Hehman E, Freeman JB: **A dynamic structure of social trait space.** *Trends in Cognitive Sciences* 2018.
- * A review of how representational similarity analysis can yield novel insights into trait impressions, including the integration of bottom-up and top-down factors.
46. Stoler RM, Hehman E, Freeman JB: **A common trait space for social cognition.** under review.
47. Tamir DI, Thornton MA, Contreras JM, Mitchell JP: **Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence.** *Proceedings of the National Academy of Sciences* 2016, **113**:194-199.

48. Jack RE, Garrod OG, Yu H, Caldara R, Schyns PG: **Facial expressions of emotion are not culturally universal.** *Proceedings of the National Academy of Sciences* 2012, **109**:7241-7244.