HHS Public Access

Author manuscript

Curr Opin Psychol. Author manuscript; available in PMC 2019 December 01.

Published in final edited form as:

Curr Opin Psychol. 2018 December ; 24: 83–91. doi:10.1016/j.copsyc.2018.10.003.

The neural representational geometry of social perception

Jonathan B. Freeman[*], **Ryan M. Stolier**, **Jeffrey A. Brooks**, and **Benjamin A. Stillerman**

New York University

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, behavioral measures, and computational modeling. Among other uses, 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.

When we encounter others, we perceive their social categories, emotional state, and even personality traits. Across social perception lies underlying representations – of social categories, emotions, traits – that in turn drive perceptual judgments and behavior. Here, we propose that important advances in social perception can be gained by triangulating on the structure of representations derived from multiple levels of analysis. Here we focus on three such levels of analysis: neuroimaging, behavioral measures, and computational 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

Computational models of social perception [1-3] assume that any given representation (e.g., the social category “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

[*]**Corresponding author**: Jonathan B. Freeman, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003.

Conflict of interest

Declarations of interest: none

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'; 4].

We can conceive of representations in social perception (e.g., social categories, emotions, traits) as points in a multidimensional space. The space can be measured using a variety of different modalities, such that the dimensions consist of neurons, fMRI voxels, or nodes in a computational model, such as a neural-network model. The specific representational structure can be estimated through patterns of activation from these different modalities, as well as through 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; 5,6] – can inform our understanding not only of how representational spaces underlying social perception manifest at the neural level, but also help address fundamental questions about how social perception is shaped by both relevant bottom-up visual and top-down social cognitive processes.

In certain neural-network models and in the actual brain, neural-representational patterns operate as ‘attractors’, such that a neural system is 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 [7]. 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 attractor dynamics have long been observed in local populations of neurons in actual cortex, and serve as an intrinsic pattern-completion process allowing neuronal patterns to serve as stored representations [7].

Computational models.

Various forms of computational models may be used to corroborate representational geometries in social perception, including neural-network models. Imagine a system of 100 neurons, with two unique patterns of activation for the emotion categories Angry and Happy; 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 Angry and Happy attractor state. Thus, at any moment, the neuronal population’s proximity to the Angry and Happy 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. The system will gravitate toward energy minima, which are the stored representational patterns, i.e., attractor states [7]. 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 Angry and Happy category representations in the system (Fig. 1).

Behavior.

To measure these attractor dynamics during social perception, researchers have recently leveraged response-directed hand movement using computer mouse-tracking [8]. 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 [8,9]. 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 [8]. Thus, 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 that representation and a perceiver's current state in higher-dimensional neural state space (Fig. 1). Numerous other behavioral measures can be used, such as ratings or response times.

Neuroimaging.

Finally, computational models and behavioral measures can be combined with neuroimaging. FMRI studies have increasingly focused on multi-voxel patterns of activity within a functional region to understand representational structure [5,11]. 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 patterns [11,12]. Recent multi-voxel pattern analyses have demonstrated reliable decoding of various kinds of representations in social perception [13]. For instance, studies have shown the fusiform gyrus (FG), a visual face-processing region [14], is involved in representing faces' gender [15], race [16,17], and emotion [18]. 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 between representations in the multidimensional space of a computational model, or physical distance within the two-dimensional space of a computer screen with mouse-tracking (or other behavioral measures), distances between representations in the multidimensional space of neural response patterns can be used to understand the structure of representations in social perception (Fig. 1). Depending on the theoretical question at hand, these neural response patterns can be localized to a specific region or distributed over a larger cortical area or even the entire brain. Moreover, using computational models of the function of different anatomical regions (e.g., the HMAX model of ventral-temporal cortex; [19]), the structure of simulated representations in a putative region can be corroborated with the structure of fMRI response patterns in that actual region.

Linking these levels of analysis – computational modeling, behavioral data, 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. 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 [13] – a process for which representational geometry has recently been useful in understanding [20] (Fig. 2).

This category competition driving perception may also be affected by top-down factors. For instance, being exposed to a face has long been known to spontaneously activate relevant gender or racial stereotypes [21,22], but recent work has suggested a more bidirectional relationship. Like other forms of perception, top-down expectations may facilitate certain perceptual interpretations [23,24], which in social perception may include stereotype-based expectations [2,13]. Over the years, stereotype effects have been documented on perceptual judgments [25-27], but it has remained less clear how such biases might manifest earlier in the perceptual process.

In one set of neuroimaging studies [10], 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 fMRI), subjective perception of faces (via mouse-tracking), and stereotype knowledge (via a stereotype content task). 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 faces' multi-voxel response patterns. For example, the more similarly the categories 'Black' and 'anger' were believed to be in terms of their stereotype knowledge 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. 3a). That these stereotypical biases on perception manifest in a visual face-processing region's representational structure suggests that social cognitive processes can impact visual representation.

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 [28,29]. Early "basic emotion" approaches emphasized six universal facial expressions of emotion, 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. Recent

work has used representational geometry to provide a comprehensive test of how inter-individual variability in conceptual knowledge may shape facial emotion perception.

In one set of studies [30], 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, controlling for faces' inherent physical similarity. 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) (Fig. 3b). 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 [31] 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 suggesting its impact on visual representation of faces. 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 work has used representational geometry to test for shared neural representations across faces and voices [32] or competing models of inferring the emotional experience of another person [33]. For instance, in response to emotional episodes, the similarity structure of multi-voxel patterns in the TPJ and dmPFC, 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 (also see Fig. 4).

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 [34,35], a process requiring minimal visual exposure [36,37]. Popular models of face-based trait impressions focus exclusively on the role of bottom-up facial features [34,38]. However, an emerging literature has noted the considerable influence of top-down cognitive processes in face impressions [6,35,39,40], for which representational geometry has recently been valuable.

One set of studies [41] tested whether individual differences in 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 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. 3c). This showed that face impressions arise not just from bottom-up features but also the conceptual understanding of what those traits mean. Although top-down factors have long been recognized as affecting personality inferences in general [42], current models of face-based trait impressions focus on a relatively fixed and universal two-dimensional structure derived from bottom-up facial features alone. These findings suggest a more dynamic multidimensional structure that varies across perceivers and contexts, not necessarily tied to any evolutionary or functional value [6]. It also empirically connects more recent face-based trait impressions research to more classic work on the conceptual shaping of personality inferences.

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 [43]. 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 [44,45], 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.

Conclusion

Across various domains of social perception, an emerging focus on representational geometry is exploring how both bottom-up facial features and a variety of top-down social cognitive factors together shape perceptions of other people. While the general correlational approach with representational geometry has permitted broad tests of how an entire system of representations in social perception may be determined by novel factors (e.g., stereotypes, emotion-concept knowledge), future research could establish more causal links through manipulating these factors. Although this work is in its early stages, linking across computational modeling, neuroimaging, and behavioral data allows researchers to use multiple levels of analysis to constrain theoretical models. 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. [PubMed: 9457773]
2. Freeman JB, Ambady N: A dynamic interactive theory of person construal. Psychological Review 2011, 118:247–279. [PubMed: 21355661]
3. Kunda Z, Thagard P: Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. Psychological Review 1996, 103:284–308.
4. Averbeck BB, Latham PE, Pouget A: Neural correlations, population coding and computation. Nature reviews neuroscience 2006, 7:358. [PubMed: 16760916]
5. Diedrichsen J, Kriegeskorte N: Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS computational biology 2017, 13:e1005508. [PubMed: 28437426]
6. Stolier RM, Hehman E, Freeman JB: A dynamic structure of social trait space. Trends in Cognitive Sciences 2018.
7. Zylberberg J, Strowbridge BW: Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory. Annual review of neuroscience 2017, 40:603–627.
8. Freeman JB: Doing psychological science by hand. Current Directions in Psychological Science 2018.
9. Freeman JB, Dale R, Farmer TA: Hand in motion reveals mind in motion. Frontiers in Psychology 2011, 2:59. [PubMed: 21687437]
10. Stolier RM, Freeman JB: Neural pattern similarity reveals the inherent intersection of social categories. Nature Neuroscience 2016, 19:795–797. [PubMed: 27135216]
11. Haynes J-D: A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. Neuron 2015, 87:257–270. [PubMed: 26182413]
12. 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. [PubMed: 20868757]
13. Freeman JB, Johnson KL: More than meets the eye: Split-second social perception. Trends in Cognitive Sciences 2016.
14. Grill-Spector K, Weiner KS, Kay K, Gomez J: The functional neuroanatomy of human face perception. Annual review of vision science 2017, 3:167–196.
15. 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. [PubMed: 19767310]
16. Contreras JM, Banaji MR, Mitchell JP: Multivoxel patterns in fusiform face area differentiate faces by sex and race. PloS one 2013, 8:e69684. [PubMed: 23936077]
17. 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. [PubMed: 22661619]
18. 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 multivoxel pattern analysis. Cortex 2015.
19. Serre T, Oliva A, Poggio T: A feedforward architecture accounts for rapid categorization. Proceedings of the national academy of sciences 2007, 104:6424–6429.
20. Stolier RM, Freeman JB: A neural mechanism of social categorization. Journal of Neuroscience 2017, 37:5711–5721. [PubMed: 28483974]
21. Macrae CN, Bodenhausen GV: Social cognition: Thinking categorically about others. Annual Review of Psychology 2000, 51:93–120.
22. Mason MF, Cloutier J, Macrae CN: On construing others: Category and stereotype activation from facial cues. Social Cognition 2006, 24:540–562.

23. O'Callaghan C, Kveraga K, Shine JM, Adams RB, Jr, Bar M: Predictions penetrate perception: Converging insights from brain, behaviour and disorder. Consciousness and cognition 2017, 47:63–74. [PubMed: 27222169]
24. Hanks TD, Summerfield C: Perceptual decision making in rodents, monkeys, and humans. Neuron 2017, 93:15–31. [PubMed: 28056343]
25. 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.
26. Bijlstra G, Holland RW, Dotsch R, Hugenberg K, Wigboldus DH: Stereotype associations and emotion recognition. Personality and Social Psychology Bulletin 2014:0146167213520458.
27. Freeman JB, Penner AM, Saperstein A, Scheutz M, Ambady N: Looking the part: Social status cues shape race perception. PLoS ONE 2011, 6:e25107. [PubMed: 21977227]
28. Barrett LF: How emotions are made: The secret life of the brain: Houghton Mifflin Harcourt; 2017.
29. Barrett LF: The theory of constructed emotion: an active inference account of interoception and categorization. Social cognitive and affective neuroscience 2017, 12:1–23. [PubMed: 27798257]
30. Brooks JA, Freeman JB: Conceptual knowledge predicts the representational structure of facial emotion perception. Nature Human Behaviour in press.
31. Brooks JA, Chikazoe J, Sadato N, Freeman JB: The neural representation of emotion perception reflects cultural and individual variability in conceptual knowledge. in prep.
32. Kuhn LK, Wydell T, Lavan N, McGettigan C, Garrido L: Similar representations of emotions across faces and voices. Emotion 2017, 17:912. [PubMed: 28252978]
33. Skerry AE, Saxe R: A common neural code for perceived and inferred emotion. Journal of Neuroscience 2014, 34:15997–16008. [PubMed: 25429141]
34. Oosterhof NN, Todorov A: The functional basis of face evaluation. Proceedings of the National Academy of Sciences 2008, 105:11087–11092.
35. 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. [PubMed: 28481616]
36. Willis J, Todorov A: First impressions: Making up your mind after a 100-ms exposure to a face. Psychological Science 2006, 17:592–598. [PubMed: 16866745]
37. Freeman JB, Stolier RM, Ingbretsen ZA, Hehman EA: Amygdala responsivity to high-level social information from unseen faces. The Journal of Neuroscience 2014, 34:10573–10581. [PubMed: 25100591]
38. Vernon RJ, Sutherland CA, Young AW, Hartley T: Modeling first impressions from highly variable facial images. Proceedings of the National Academy of Sciences 2014, 111:E3353–E3361.
39. Sutherland CA, Young AW, Mootz CA, Oldmeadow JA: Face gender and stereotypicality influence facial trait evaluation: Counter- stereotypical female faces are negatively evaluated. British Journal of Psychology 2015, 106:186–208. [PubMed: 25168952]
40. Sutherland CA, Oldmeadow JA, Young AW: Integrating social and facial models of person perception: Converging and diverging dimensions. Cognition 2016, 157:257–267. [PubMed: 27689511]
41. Stolier RM, Hehman E, Keller M, Walker M, Freeman JB: The conceptual structure of face impressions. PNAS invited revision.
42. Schneider DJ: Implicit personality theory: A review. Psychological bulletin 1973, 79:294. [PubMed: 4574836]
43. Stolier RM, Hehman E, Freeman JB: A common trait space for social cognition. under review.
44. 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.
45. Thornton MA, Mitchell JP: Theories of person perception predict patterns of neural activity during mentalizing. Cerebral cortex 2017:1–16. [PubMed: 28365777]

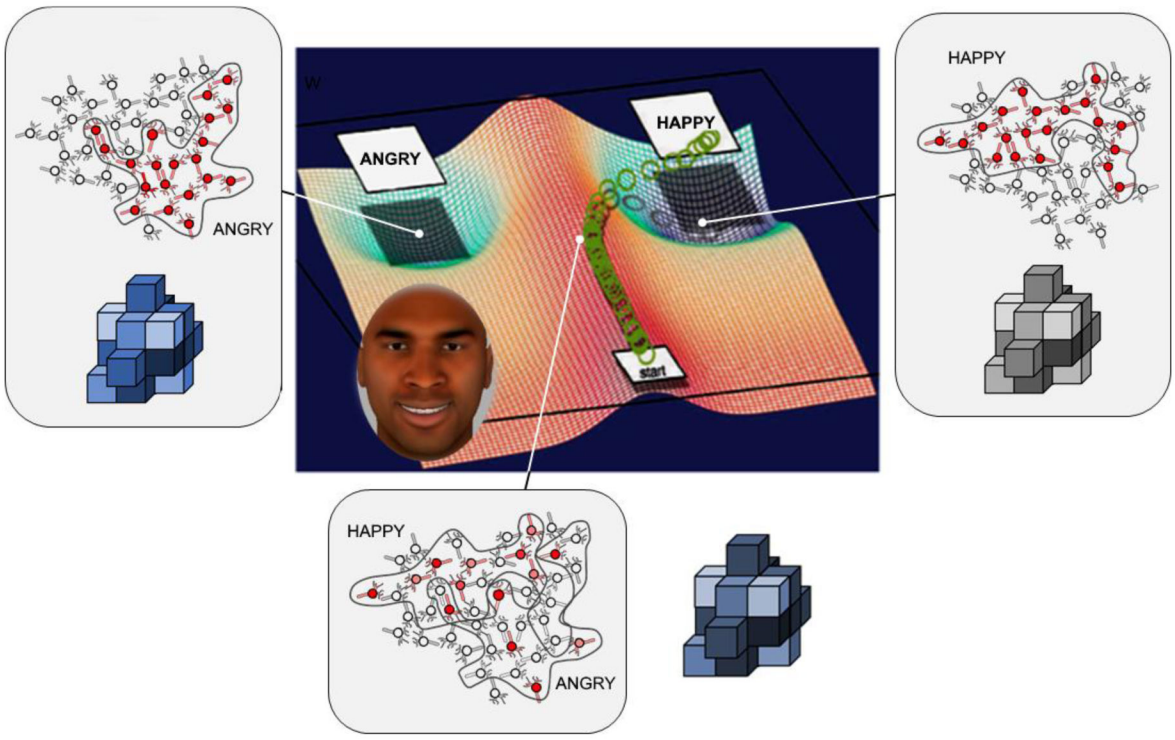

Figure 1. The link between modeling, mouse-tracking, and multivariate fMRI.

An illustration of the link between the three modalities is depicted using an example of stereotype influences on emotion perception based on [10]. Happy and Angry categories are each associated with a unique pattern across a neuronal population, with certain neurons highly active during a 'Happy state and other neurons highly active during an 'Angry' 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 fMRI. 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 Happy and Angry response locations and 'Happy and 'Angry' 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., Happy or Angry category, just as the hand settles into one of the response locations. Due to automatic stereotype-based expectations linking Black people to hostility, for a happy Black face, during the perceptual process (e.g., mid-trajectory) the neuronal pattern would approximate the Angry pattern to a greater extent and the hand would be more attracted toward the Angry response (e.g., relative to a happy White face). Because the multi-voxel pattern in response to such a face would reflect an average over this time period (as fMRI's temporal sensitivity is limited), it would exhibit a degree of greater pattern-similarity to the Angry category (e.g., relative to a happy White face), as

shown in previous neuroimaging work [10]. 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.

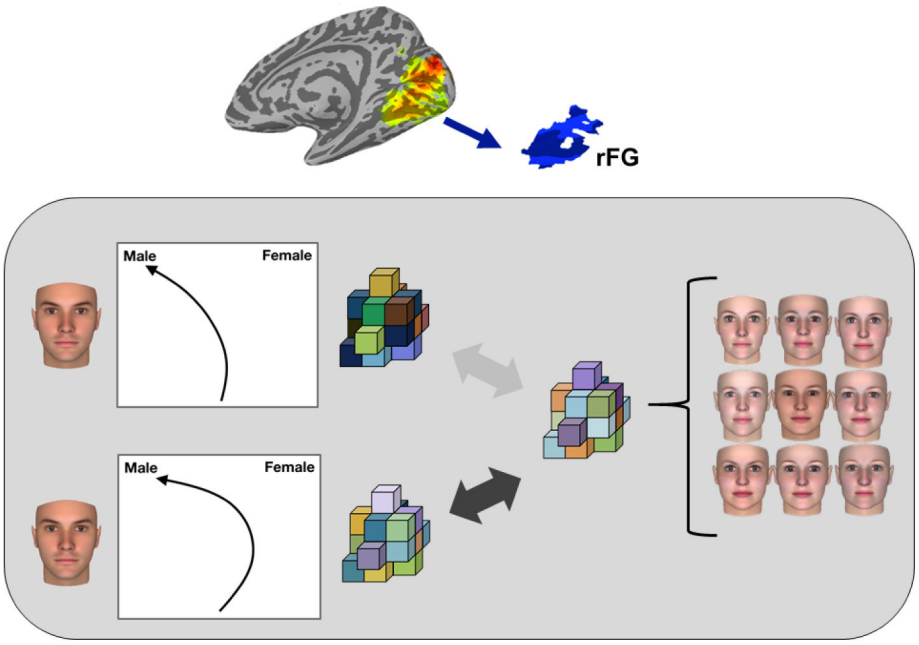

Figure 2. Co-activation of representations in social categorization.

A recent study [20] 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, suggesting 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 [20]. 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.

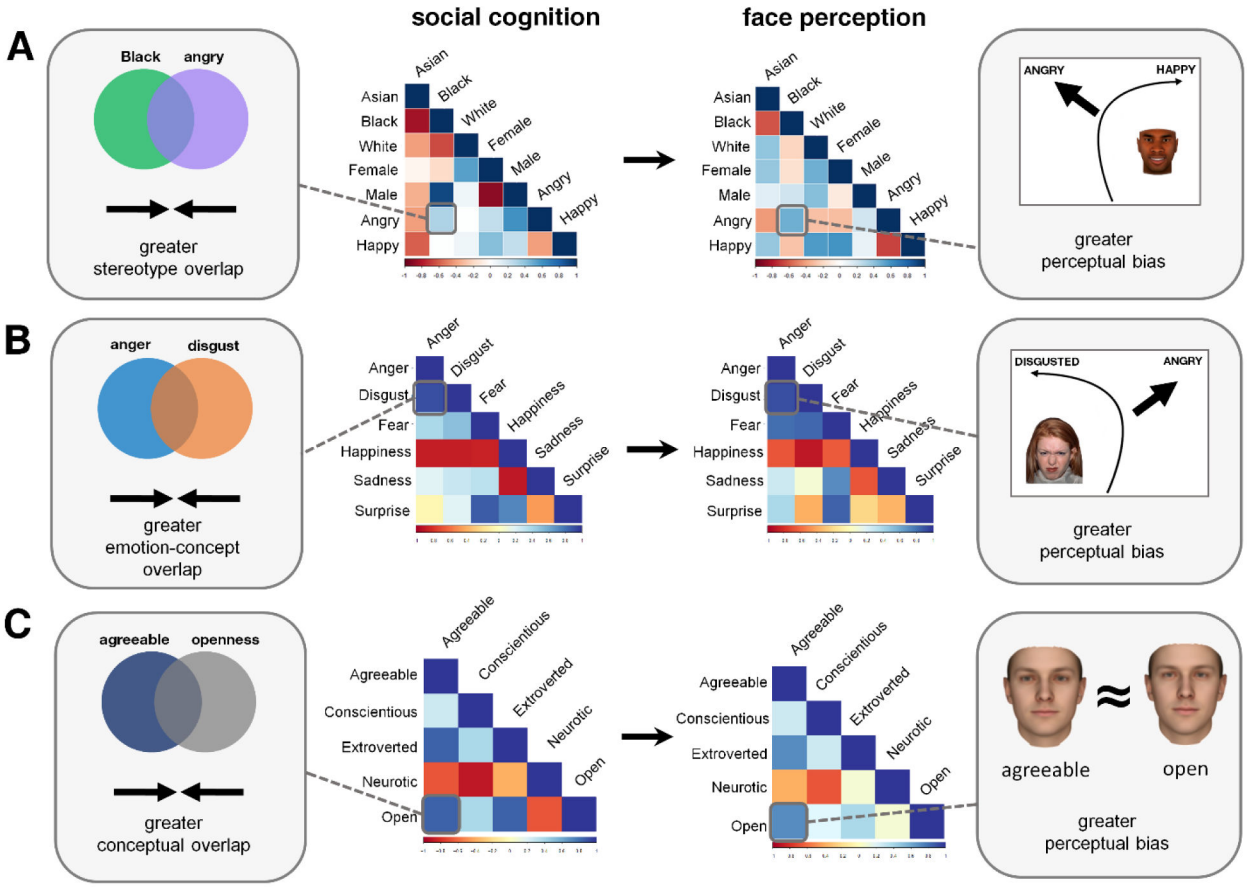

Figure 3. Representational similarity analysis (RSA) in social perception.

In RSA, representational dissimilarity matrices (RDMs) comprise all pairwise similarities/dissimilarities (see Fig. 4A) 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 [10]. **(B)** Participants' emotion-concept RDM (emotion concept 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 [30], which was also reflected in FG pattern structure [31]. **(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 [41].

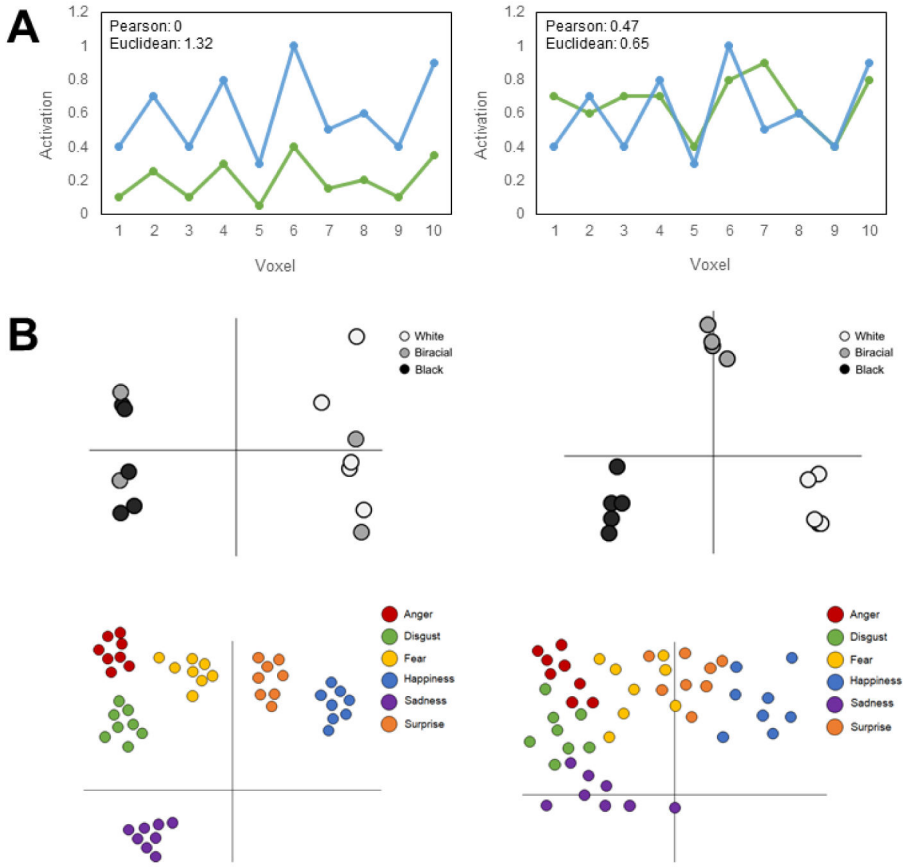

Fig. 4. Representational similarity and geometry.

(A) Representational dissimilarity matrices (RDMs), as in Fig. 3, comprise dissimilarities for all pairs of conditions. For simple cases, these dissimilarities – i.e., the distances between pairs of representations – can be derived by single values (e.g., behavioral similarity rating, mouse-trajectory deviation). In most cases, however, data are multivariate, and pairwise dissimilarities must be computed using distance metrics such as Pearson correlation distance or Euclidean distance. Consider the activation pattern of a brain region with 10 voxels. Pearson distance normalizes vector magnitude and scale, and thus in the case of multi-voxel patterns removes any differences in overall activation level and variability, leaving only the relative shape of the pattern that is typically of interest. Euclidean distance, on the other hand, combines sensitivity to relative pattern shape with absolute differences in magnitude and scale. On the left, for the two conditions' multi-voxel patterns, Pearson distance is zero, because normalizing for magnitude and scale differences, the pattern is identical; Euclidean distance, however, is sensitive to these differences (i.e., how much higher one pattern is than another). On the right, the relative shape of the two conditions' multi-voxel patterns are much less congruent, leading Pearson distance to be higher than on the left; Euclidean distance is lower than on the left, because incorporating magnitude and scale, the patterns are now more similar. Consider if the vector comprised 10 items on a stereotype survey (e.g., aggressive, communal), rather than 10 voxels, and the conditions were two social groups, such absolute differences in magnitude and scale may be important to capture (e.g., that one group is judged overall more aggressive and less communal than another). **(B)** 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. Top panel: When presented with White, Biracial, and Black faces, one 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). Bottom panel: When presented with emotional faces, certain perceivers may have a 6-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. Thus, examining representational spaces in social perception may reveal important differences in the perceptual organization of different brain regions or individual perceivers.